PB 211 265

COLLABORATIVE STUDY
of
REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

Herbert C. McKee
Ralph E. Childers

Contract CPA 70-40
SwRI Project 01-2811

Prepared for
Office of Measurement Standardization
Division of Chemistry and Physics
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, N. C. 27709

May 1972

PROPERTY OF ... MEASUREMENT ...

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

SOUTHWEST RESEARCH INSTITUTE
SAN ANTONIO HOUSTON

BIBLIOGRAPHIC DATA SHEET	1. Report No.	2.	3. Recipient's Accession No. PB-211 265
4. Title and Subtitle Collaborative Study of Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry)			5. Report Date May 1972 6.
7. Author(s) Herbert C. McKee and Ralph E. Childers			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Southwest Research Institute 3600 South Yoakum Boulevard Houston, Texas 77006			10. Project/Task/Work Unit No. SwRI 01-2811 11. Contract/Grant No. CPA 70-40
12. Sponsoring Organization Name and Address ENVIRONMENTAL PROTECTION AGENCY National Environmental Research Center, Division of Chemistry and Physics, Office of Measurement Standardization Research Triangle Park, North Carolina 27711			13. Type of Report & Period Covered 14.
15. Supplementary Notes			

16. Abstracts. *Information obtained in the evaluation and collaborative testing of a reference method for measuring the carbon monoxide content of the atmosphere is presented. The method is based on the infrared absorption characteristics of carbon monoxide, using an instrument calibrated with gas mixtures containing known concentrations of carbon monoxide. The method as published in the appended Federal Register article was tested by means of a collaborative test involving a total of 16 laboratories. The test involved the analysis of both dry and humidified mixtures of carbon monoxide and air over the concentration range from 0 to 60 mg/m^3. A statistical analysis of the data of 15 laboratories provided the following results: the checking limit for duplicates is 0.5 mg/m^3; the repeatability is 1.6 mg/m^3; the reproducibility varies nonlinearly with concentration with a minimum of 2.3 mg/m^3 at a concentration of 20 mg/m^3 and ranges as high as 4.3 mg/m^3 in the concentration range of 0 to 60 mg/m^3; the minimum detectable sensitivity is 0.3 mg/m^3; the compensation for water vapor interference is satisfactory for drying agents and refrigeration methods. The use of narrow-band optical filters alone may not provide adequate compensation; and the accuracy is totally dependent upon the availability of dependable calibration standards; the method produces results, on the average, 2.5 percent high. Other results with respect to the quality of calibration standards and the minimum number of samples required to establish validity of results within stated limits are also presented.*

17. Key Words and Document Analysis. 17a. Descriptors

Air Pollution
Carbon Monoxide
Infrared Spectroscopy
Calibrating
Laboratories
Statistical Analysis
Water
Vapors

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group 13 B 11

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Unlimited	UNCLASSIFIED	48
	20. Security Class (This Page) UNCLASSIFIED	22. Price $3.00 - .75 MF

Errata
Appendix A

Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry)

Page A-1, Section 1.1, lines 4 and 5
delete "split into parallel beams and"

COLLABORATIVE STUDY
of
REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

Herbert C. McKee
Ralph E. Childers

Contract CPA 70-40
SwRI Project 01-2811

Prepared for
Office of Measurement Standardization
Division of Chemistry and Physics
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, N. C. 27709

May 1972

Approved

Herbert C. McKee

Herbert C. McKee
Assistant Director
Department of Chemistry
and Chemical Engineering

SOUTHWEST RESEARCH INSTITUTE - HOUSTON
3600 SOUTH YOAKUM BOULEVARD, HOUSTON, TEXAS 77006

SUMMARY AND CONCLUSIONS

This report presents information obtained in the evaluation and collaborative testing of a reference method for measuring the carbon monoxide content of the atmosphere.

This method was published by the Environmental Protection Agency in the *Federal Register*, April 30, 1971, as the reference method to be used in connection with Federal ambient air quality standards for carbon monoxide. Following minor editorial changes, the method was republished in the *Federal Register*, November 25, 1971. The former publication is reproduced as Appendix A of this report.

The method is based on the infrared absorption characteristics of carbon monoxide, using an instrument calibrated with gas mixtures containing known concentrations of carbon monoxide. A similar method based on the same principle has been published by the Intersociety Committee as Tentative Method 42101-04-69T in **Health Laboratory Science**, January 1970, Part Two, pp 81-86.

The method published in the *Federal Register* was tested, as a part of this program, by means of a collaborative test involving a total of 16 laboratories. The test involved the analysis of both dry and humidified mixtures of carbon monoxide and air over the concentration range from 0 to 60 mg/m^3. A statistical analysis of the data of 15 laboratories provided the following results:

- The checking limit for duplicates is 0.5 mg/m^3.
- The repeatability is 1.6 mg/m^3.
- The reproducibility varies nonlinearly with concentration with a minimum of 2.3 mg/m^3 at a concentration of 20 mg/m^3 and ranges as high as 4.3 mg/m^3 in the concentration range of 0 to 60 mg/m^3.
- The minimum detectable sensitivity is 0.3 mg/m^3.
- The compensation for water vapor interference is satisfactory for drying agents and refrigeration methods. The use of narrow-band optical filters alone may not provide adequate compensation.
- The accuracy is totally dependent upon the availability of dependable calibration standards. Based on the results of this collaborative study, the method produces results, on the average, 2.5 percent high.

In addition, this report presents other results with respect to the quality of calibration standards and the minimum number of samples required to establish validity of results within stated limits.

ACKNOWLEDGEMENT

The authors wish to express appreciation to the Project Officer, Mr. Thomas W. Stanley, and staff member, Mr. John H. Margeson, of the Office of Measurement Standardization, for assistance in the planning and execution of the collaborative study.

The assistance and advice of Scott Research Laboratories, Plumsteadville, Pennsylvania, who prepared and analyzed the test gases, is acknowledged.

The assistance and cooperation of the participating laboratories is also acknowledged with sincere appreciation for the voluntary efforts of the staff members who represented each organization. The representatives and organizations participating in one or more phases of the collaborative test program were as follows:

Name	Organization
Richard S. Brief R.G. Confer	Esso Research and Engineering Company Linden, New Jersey
Franz J. Burmann Jack A. Bowen	Environmental Protection Agency Durham, North Carolina
Charles A. Cody	Southwest Research Institute Houston, Texas
Ronald P. Dubin Richard Wonderlick	Bureau of Air Pollution Control Pennsylvania Department of Environmental Resources Harrisburg, Pennsylvania
Milton Feldstein	Bay Area Air Pollution Control District San Francisco, California
Judith Garelick Diane Berkel	Bureau of Air Pollution Control Nassau Department of Health Mineola, New York
Kenneth T. Irwin	Jefferson County, Kentucky Air Pollution Control District Louisville, Kentucky
Norman J. Lewis	Division of Environmental Quality New Jersey State Department of Health Trenton, New Jersey

Name	Organization
Peter K. Mueller S.G. Kerns	Air and Industrial Hygiene Laboratory California State Department of Public Health Berkeley, California
Robert E. Pattison F.D. Olmstead	Air Pollution Control Laboratory Canton City Health Department Canton, Ohio
Hisham M. Sa'aid Roger B. McCann	Air Quality Section Kentucky Air Pollution Control Commission Frankfort, Kentucky
R.K. Stevens Dwight A. Clay	Environmental Protection Agency Research Triangle Park, North Carolina
Bill Stewart J.S. Payne	Air Pollution Control Texas State Department of Health Austin, Texas
Philip S. Tow	Sacramento County, California Air Pollution Control District Sacramento, California
Alvin L. Vander Kolk Ken Smith	Standards and Analysis Sections Michigan Department of Public Health Lansing, Michigan
Grant S. Winn Carl E. Kerr	Air Quality Section Utah State Division of Health Salt Lake City, Utah

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	COLLABORATIVE TESTING OF THE METHOD	1
	A. Furnishing Test Samples and Calibration Gases	2
	B. Selection of Collaborators	3
	C. Collabortive Test Procedure	4
III.	STATISTICAL DESIGN AND ANALYSIS	4
	A. Summary of Design	4
	B. Summary of Results	5
	LIST OF REFERENCES	7

LIST OF ILLUSTRATIONS

Figure		Page
1	Repeatability and Reproducibility Versus Concentration	6

LIST OF TABLES

Table		Page
1	Reference Values for Carbon Monoxide Test Concentrations Used in Collaborative Test, Parts per Million	3

Preceding page blank

I. INTRODUCTION

While carbon monoxide has received less attention than some other contaminants, it is found in many of the urban areas of the world. Many different sources of carbon monoxide exist in a typical city, but by far the predominant source is motor vehicles. As recent control measures reduce the emissions from vehicles, other sources such as incinerators and various industrial operations will represent a larger percentage of the total.

Carbon monoxide has long been known to be toxic at high concentrations, producing illness and eventually death. At the lower levels of concentration found in many urban atmospheres, carbon monoxide may act to impair various bodily functions, although the exact exposure conditions required to produce such effects have not been definitely established.

Unlike most atmospheric contaminants, attempts to measure carbon monoxide by a direct chemical method have met with very limited success. For industrial hygiene purposes, combustion processes and colored indicator tubes have been used satisfactorily. To measure the lower concentrations of interest in urban air pollution, however, the only satisfactory method which has received widespread use is based on infrared absorption. Commercial instruments based on this principle have been available for many years. The method involves detecting the difference in absorption of infrared energy of the atmosphere being tested in a sample cell and a non-absorbing gas in a reference cell. The difference is sensed by selective detectors, sensitive only to carbon monoxide, and amplified to provide an output signal. The resulting signal is then used to operate a recorder which provides a continuous record of carbon monoxide levels over a period of time. Under normal atmospheric conditions, the only major interference with this method is water vapor, which can be overcome through the use of drying agents or other measures as discussed subsequently.

In order to obtain reliable data in measuring carbon monoxide and other atmospheric contaminants, the Environmental Protection Agency (EPA) Office of Measurement Standardization (OMS) has been working for some time to develop standard methods which could be used by all persons making air quality measurements. Following the development of a tentative standard method, the final step in the standardization process is a collaborative test, or interlaboratory comparison, of the proposed standard method. This procedure, also called "round-robin testing," has been used to evaluate many different methods of measurement in such diverse fields as water chemistry, metallurgy, paint and surface coatings, food and related products, and many others. A test of this nature by a representative group of laboratories is the only way that the statistical limits of error inherent in any method can be determined with sufficient confidence.

This report presents the results of a collaborative test of the carbon monoxide method conducted by Southwest Research Institute and the Office of Measurement Standardization, together with the statistical analysis of the data obtained. In this collaborative test, standard samples contained in high-pressure cylinders were prepared and carefully analyzed to determine exact concentration. These cylinders were then distributed to a representative group of laboratories who participated in the test on a voluntary basis. These samples were analyzed according to the standard procedure as outlined in the tentative method, after which the gas cylinders were returned to the supplier for reanalysis to again check the concentration levels. The results of the collaborative test were then analyzed statistically to determine the accuracy and precision of the proposed method.

II. COLLABORATIVE TESTING OF THE METHOD

An important step in the standardization of any method of measurement is the collaborative testing of a proposed method to determine, on a statistical basis, the limits of error which can be expected when the method is used by a typical group of

Preceding page blank

investigators. The collaborative, or interlaboratory, test of a method is an indispensable part[1]* of the development and standardization of an analytical procedure to insure that (1) the procedure is clear and complete and that (2) the procedure does give results with precision and accuracy in accord with those claimed for the method. Among other organizations, the Association of Official Analytical Chemists (AOAC) and the American Society for Testing and Materials (ASTM) have been active in the field of collaborative testing and have published guidelines of the proper procedure for conducting collaborative tests and evaluating the data obtained.[2-4] Publications of both organizations were used extensively in planning and conducting the collaborative tests of this method to measure carbon monoxide.

After the evaluation of various methods for preparing test samples, a detailed collaborative test was undertaken to obtain the necessary data to make a statistical evaluation of the method. This section of the report describes the various phases of the test plan that was developed.

A. Furnishing Test Samples and Calibration Gases

Many air contaminants must be measured at concentrations in the fractional parts-per-million range, and the use of test atmospheres in high-pressure cylinders is not feasible at these low levels due to reaction or adsorption effects which make it impossible to maintain an accurately controlled test concentration. This is not true with carbon monoxide, which is only of concern at levels in the parts-per-million range. At these higher levels, and with proper precautions, cylinder gas samples can be used with a reasonable degree of confidence in the stability of the samples. The stability should be checked by periodic reanalysis where possible.

Test gases for this collaborative test were obtained from Scott Research Laboratories, an organization with wide experience in the generation, control, and analysis of various gases for experimental purposes. For each test concentration, a large master cylinder containing carbon monoxide in dry synthetic air was prepared and analyzed accurately by gas-solid chromatography using helium ionization detection. The chromatograph was calibrated with primary gravimetric gaseous standards prepared in glass. From these master cylinders, smaller cylinders were filled and individually analyzed by the same method. The cylinders used had a chromium-molybdenum alloy inside surface of low iron content to minimize the loss of carbon monoxide which has been reported to be caused by the formation of iron carbonyl.[5] The master cylinders were retained and the smaller cylinders sent to the collaborative test participants.

At the conclusion of their analyses, the participants returned the cylinders to Scott Research Laboratories, who then reanalyzed the contents of each cylinder having sufficient residual pressure. The date of the first analysis was July 2, 1971; the date of the reanalysis was January 21, 1972–203 days later. The results are shown in Table I. They are shown in parts per million as reported (divide by 0.873 to convert to milligrams per cubic meter). The results have not been converted in Table I in order to eliminate misleading comparisons because of round-off errors due to conversion. The converted values may be seen in Table C-II of Appendix C.

Agreement between first and final analyses was good for all but 3 or 4 of the 48 cylinders used. Most collaborators completed their work within 30 days of the first analysis; all work was complete within 80 days. Therefore, if any corrections were to be applied, the result would be much closer to the first analysis. No corrections were applied and the first analyses were used as the reference values.

Since the infrared instrument produces a relative measurement, calibration with standard gases is necessary in order to convert this measurement to a measured concentration as described in the method (Appendix A). This is done by the use of calibration

*Superscript numbers in parentheses refer to the List of References.

TABLE I. REFERENCE VALUES FOR CARBON MONOXIDE TEST CONCENTRATIONS USED IN COLLABORATIVE TEST, PARTS PER MILLION

Assignee	Cylinder Number	Initial Analysis	Final Analysis	Change
Master	W-18156	7.50	7.47	−0.03
220	C-658	7.29	7.27	−0.02
222	C-657	7.47	–	–
253	C-656	7.40	–	–
270	C-655	7.33	7.21	−0.12
310	C-654	7.43	7.53	0.10
311	C-653	7.49	–	–
370	C-652	7.48	7.52	0.04
375	C-651	7.20	6.54	−0.66
540	C-650	7.42	7.46	0.04
571	C-649	7.45	7.44	−0.01
780	C-648	7.35	7.33	−0.02
799	C-644	7.36	7.33	−0.03
860	C-647	7.13	5.50	−1.63
920	C-646	7.48	7.28	−0.20
923	C-645	7.48	7.57	0.09
927	C-641	7.36	7.22	−0.14
Master	W-138035	25.5	25.2	−0.3
220	C-674	26.1	26.8	0.7
222	C-673	26.4	–	–
253	C-672	26.3	–	–
270	C-671	26.1	26.2	0.1
310	C-670	26.0	26.0	0.0
311	C-669	26.0	26.3	0.3
370	C-668	26.1	–	–
375	C-667	26.4	26.2	−0.2
540	C-666	26.1	26.1	0.0
571	C-665	26.2	26.4	0.2
780	C-664	26.0	26.0	0.0
799	C-659	26.4	26.2	−0.2
860	C-663	26.1	26.4	0.3
920	C-662	26.0	25.9	−0.1
923	C-661	26.3	25.9	−0.4
927	C-660	26.3	26.3	0.0
Master	W-138036	45.5	45.5	0.0
220	C-690	45.9	45.7	−0.2
222	C-689	45.5	–	–
253	C-688	45.7	–	–
270	C-687	45.6	45.7	0.1
310	C-686	45.6	45.7	0.1
311	C-685	45.9	45.6	−0.3
370	C-684	45.5	–	–
375	C-683	45.7	45.6	−0.1
540	C-682	45.7	–	–
571	C-681	45.7	45.7	0.0
780	C-680	45.6	45.7	0.1
799	C-675	45.7	45.7	0.0
860	C-678	45.7	45.6	−0.1
920	C-677	45.7	45.8	0.1
923	C-676	45.7	45.4	−0.3
927	C-679	45.7	45.2	−0.5

Source: Scott Research Laboratories

gases representing 20, 40, 60, and 80 percent of the range of the instrument. Such calibration gases are available from a number of commercial suppliers, and the participants were instructed to obtain the necessary calibration gases from their usual sources.

Because of this calibration procedure, the accuracy of the method is completely dependent on the accuracy of the calibration gases used. For this reason, all collaborators were instructed to take all possible precautions to obtain calibration gases of sufficient accuracy and to safeguard these materials from contamination or deterioration in storage or use.

B. Selection of Collaborators

If a collaborative test is to achieve the desired objective, it is desirable that the participants in the test be representative of the large group that will ultimately use the method being tested. Since air pollution measurements are of interest to many different groups, it was desirable to include in the group of collaborators a variety of governmental agencies, universities, industrial laboratories, and others. The final selection of participants included two from federal laboratories, twelve from state and local air pollution control agencies, one from industry, and one from a research institution. A complete list of the participants and their affiliation is given in the acknowledgement.

Even more important than the type of laboratory is the degree of skill and experience of the persons who participated. Each laboratory was asked to assign a person to this test who had previous experience with the infrared method for measuring carbon monoxide and was competent in carrying out measurements by this method. This was done because the emphasis was upon the capabilities of the method rather than the performance of the laboratories. Each laboratory had previous experience in the use of the method and thus possessed a satisfactory infrared instrument and the necessary equipment for laboratory processing of samples and calibration gases.

For purposes of familiarization, each participant was furnished a standard test sample for analysis prior to the actual collaborative test. Results from

these preliminary runs were used as an approximate check on the experience and skill of each participant, with the intention of eliminating any whose results were grossly in error, thus indicating a lack of familiarity or experience with the method. No such elimination was necessary and, therefore, all of the participants originally selected were used in the actual collaborative test which followed.

C. Collaborative Test Procedure

After the preliminary familiarization samples were analyzed and the results obtained, test samples for the actual collaborative test were distributed to each participant. The national primary and secondary air quality standard for carbon monoxide is 10 mg/m^3 for 8 hr or 40 mg/m^3 as a maximum 1-hr concentration (both to be exceeded not more than once per year). Therefore, test concentrations were selected to indicate the variability of the method within these ranges. This led to the selection of 8, 30, and 53 mg/m^3 as test concentrations for purposes of collaborative testing.

In addition to examining the effects of concentration on precision and accuracy, it was necessary to realistically evaluate the effects of humidity on the analysis. Therefore, in addition to analyzing the dry test gases, each was analyzed after humidification. The test gases were essentially saturated by passing them through a midget impinger containing 15 mℓ distilled water. Losses of carbon monoxide due to absorption are negligible.

In order to estimate other random effects, each of the three concentrations was analyzed in triplicate on each of 3 days under both dry and humid conditions. This resulted in a total of 810 separate determinations–54 by each reporting laboratory. Section I-B of Appendix B contains a more detailed discussion of the experiment design, and Figure B-1 graphically shows the design.

The results of this test series were then used for detailed statistical analysis, which is described in detail in Appendix B and summarized in the next section.

III. STATISTICAL DESIGN AND ANALYSIS

Several fundamental requirements must be met in order to provide the maximum reliability of the collaborative test. First, the conditions of the test must be representative of a specified population; each factor involved must be a representative sample of a population about which inferences are to be drawn. Second, the collaborative test must be unbiased; precautions must be taken to avoid the introduction of any bias in the collaborative test procedure. It is important that the collaborators assume a responsibility to try to eliminate any bias by carefully following the instructions of the collaborative procedure and the method. Every detail is important and even the slightest departure from the specified procedures may bias the results. Third, the results of the collaborative test must be reproducible; that is, the conditions for the test should be such that similar results would be obtained if the collaborative test were repeated. The fourth requirement involves the scope of the test; the materials and conditions for which the analytical method was designed must be included in the test. Finally, the collaborative test must be practical and economically feasible. Since funds and facilities are never available for an unlimited testing program, it is necessary to accept less than the ideal testing procedures in order to accomplish the program. Thus, fundamental requirements may not be completely fulfilled, since any practical compromise introduces limitations on the inferences that can be drawn. If pursued too far, compromises from practical considerations may render the collaborative test useless.

Appendix B contains the complete and detailed description of the design and analysis of the formal collaborative test. The results of Appendix B are summarized in this section.

A. Summary of Design

The primary purpose was to establish the reliability of the method in terms of its precision and accuracy. More emphasis was placed on the quality of the method when properly used than upon the

performance of the laboratories. At the same time, it was necessary to retrieve information which would allow the investigation of other aspects of the method; therefore, intermediate data were obtained relating to calibration curves.

The statistical planning of a program is limited in scope and depends upon what information is desired. The scope is limited by what a collaborating laboratory can conveniently and economically accomplish, as well as by the number of collaborators that can be accommodated. Under these limitations, it was possible to examine the effects of laboratories, concentrations, and days upon the precision of the method in addition to estimating the replication error.

Of the 16 laboratories that took part in the test program, 15 satisfactorily completed the test. These laboratories constitute a random sample of a rather large population of experienced laboratories. Three different concentrations were analyzed by each laboratory. The concentrations were nominally 8, 30, and 53 mg/m^3. Each of the three concentrations was analyzed both dry and humidified, in triplicate, on each of 3 separate days using independently prepared calibration curves. This procedure resulted in a total of 810 individual determinations.

The collaborative test was designed to allow the analysis of the results using the most efficient statistical methods available. The experiment was designed so that the linear model analysis[3,6-8] could be used. This analysis, as well as tests for outlying observations, is described in Appendix B.

B. Summary of Results

1. Procedural Errors

Since the emphasis was upon the quality of the method and not upon the performance of the laboratories, all arithmetic errors were corrected, and the arithmetic error problem was evaluated qualitatively. Few instances of errors in arithmetic operations were noted. The method is relatively simple and, consequently, is not vulnerable to arithmetic and procedural errors.

2. Precision Between Replicates

The replication error (see Section II-A of Appendix B for detailed definition) was shown to be independent of concentration and humidity. The standard deviation for variation between replicates is equal to 0.17 mg/m^3. Replication will not materially assist in increasing the precision of the method, and will, in general, be a waste of time and effort; however, replicates are often advisable to avoid gross errors.

The checking limit for duplicates is 0.5 mg/m^3; therefore, two replicates differing by more than this amount should be considered suspect. Section II-A of Appendix B contains more details regarding the replication error.

3. Humidity Effects

The humidity has no measurable effect upon the precision or accuracy when drying or refrigeration methods are used (see Section 3.1 of the method in Appendix A). No data are available for the saturation method. Optical filters alone do not appear to be adequate; however, this conclusion is based on very limited data. Section II-B of Appendix B contains further details regarding humidity effects.

4. Precision Between Days

The standard deviation for variation between days for the same sample includes both the replication error and a component for between-days variation and is equal to 0.47 mg/m^3. Two test results on the same sample on different days by the same laboratory should not differ by more than 1.3 mg/m^3.

The standard deviation for variation between days for different but similar samples includes an additional term to account for heterogeneity between samples. The corresponding standard deviation is 0.57 mg/m^3. If the test results on each of these samples differ by less than 1.6 mg/m^3 – the repeatability of the method--there is no reason to believe there is any real difference between them.

Section III-A-2 of Appendix B presents more details regarding precision between days; in particular, a comparison of the means of different populations each analyzed by the same laboratory.

5. Precision Between Laboratories

The standard deviation for variation between laboratories includes terms representing additional, more complex, effects and is equal to the square root of $V_j(y)$ where $V_j(y)$ is given by

$$V_j(y) = 0.001007x_j^2 - 0.0393x_j + 1.10$$

where the subscript j is attached to signify the dependence upon the concentration x_j which is the independent variable.

Two test results on the same sample should agree within the reproducibility which is shown plotted versus concentration in Figure 1 along with the repeatability for comparison. If the test results on two different samples differ by less than the reproducibility, there is no reason to believe there is any real difference between them.

Section III-A-3 of Appendix B includes more details regarding the precision between laboratories.

Various statistical methods are available for the comparison of means or the comparison of a mean and a fixed value.[9-11] These methods are straightforward and are applied independently of the results of this study. That is, whether or not a mean is

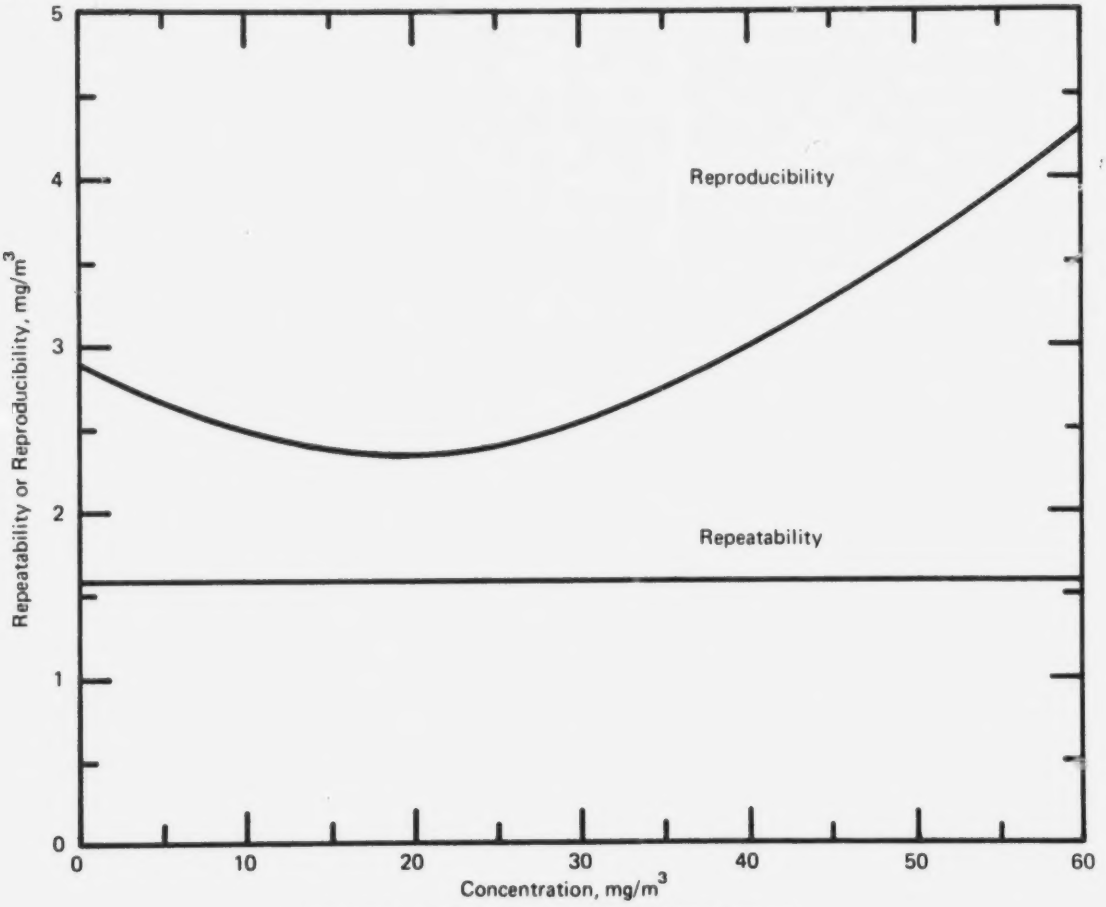

FIGURE 1. REPEATABILITY AND REPRODUCIBILITY VERSUS CONCENTRATION

significantly different from some fixed value is dependent upon the actual standard deviation of the sample population. The variance of the sample population includes both the variance of the true values and the variance due to the measurement method. A limiting case is discussed in Appendix B under the assumption that all variation is due to the measurement method. The case is an extremely unlikely, if not impossible, situation; however, a certain amount of guidance can be obtained in terms of the numbers of observations required to provide a specified degree of agreement. These numbers are sufficient only to compensate for the variation of the method. An additional quantity, dependent on the variation in the true values, will always be required. Interested readers may refer to Figures B-4 and B-5 and the respective discussions in Section III-A-3 of Appendix B where two illustrative examples are given.

6. Accuracy of the Method

There is a statistically significant bias in the method based upon the results of this collaborative test. The practical significance must be based upon other criteria.

There is an approximately linear relationship with the tendency for results to be, on the average, 2.5 percent high. (See Figure B-6 in Appendix B for a graphic illustration.) Since the method uses the same type materials for calibration as were used for reference samples in this test, there remains little doubt that the inaccuracy results almost entirely from the use of calibration gases which exhibit significant variation with respect to their specified content. Since results tend to be high, the calibration gases must have a tendency to be correspondingly low.

It cannot be overemphasized that the accuracy of the method is almost totally dependent upon the availability of sufficiently accurate calibration standards.

Section III-B of Appendix B contains further details regarding the accuracy of the method and an examination of the quality of calibration gases.

7. Minimum Detectable Sensitivity

The minimum detectable sensitivity is defined as "the smallest amount of input concentration that can be detected as the concentration approaches zero" (see Addenda B of the method in Appendix A). The best estimate for this parameter is that based on two standard deviations (replication error); therefore, the minimum detectable sensitivity may be taken to be 0.3 mg/m^3. Obviously, it is also affected by other criteria such as chart range and dimensions, recorder performance, and instrument response. These characteristics varied widely in the collaborative test.

LIST OF REFERENCES

1. Youden, W.J., "The Collaborative Test," *Journal of the AOAC*, Vol 46, No. 1, pp 55-62 (1963).

2. **Handbook of the AOAC**, Second Edition, October 1, 1966.

3. ASTM Manual for Conducting an Interlaboratory Study of a Test Method, ASTM STP No. 335, *Am. Soc. Testing & Mats.* (1963).

4. **1971 Annual Book of ASTM Standards,** Part 30, Recommended Practice for Developing Precision Data on ASTM Methods for Analysis and Testing of Industrial Chemicals, ASTM Designation:E180-67, pp 403-422.

5. Westberg, Karl; Cohen, Norman; and Wilson, K.W.: "Carbon Monoxide: Its Role in Photochemical Smog Formation," *Science*, Vol 171, No. 3975, pp 1013-1015 (March 12, 1971).

6. Mandel, John, **The Statistical Analysis of Experimental Data**, John Wiley & Sons, New York, Chapter 13, pp 312-362 (1964).

7. Mandel, J., "The Measuring Process," *Technometrics*, 1, pp 251-267 (1959).

8. Mandel, J., and Lashof, T.W., "The Interlaboratory Evaluation of Testing Methods," *ASTM Bulletin*, No. 239, pp 53-61 (1959).

9. Dixon, Wilfred J., and Massey, Frank J., Jr., **Introduction to Statistical Analysis**, McGraw-Hill Book Company, Inc., New York, Chapter 9, pp 112-129 (1957).

10. Duncan, Acheson J., **Quality Control and Industrial Statistics**, Third Edition, Richard D. Irwin, Inc., Homewood, Illinois, Chapters XXV and XXVI, pp 473-521 (1965).

11. Bennett, Carl A., and Franklin, Norman L., **Statistical Analysis in Chemistry and the Chemical Industry**, John Wiley and Sons, New York, Chapter 5, pp 149-164 (1954).

APPENDIX A

REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

Reproduced from Appendix C, "National Primary and Secondary Ambient Air Quality Standards," *Federal Register*, Vol 36, No. 84, Part II, Friday, April 30, 1971

APPENDIX C—REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

1. *Principle and Applicability.*

1.1 This method is based on the absorption of infrared radiation by carbon monoxide. Energy from a source emitting radiation in the infrared region is split into parallel beams and directed through reference and sample cells. Both beams pass into matched cells, each containing a selective detector and CO. The CO in the cells absorb infrared radiation only at its characteristic frequencies and the detector is sensitive to those frequencies. With a nonabsorbing gas in the reference cell, and with no CO in the sample cell, the signals from both detectors are balanced electronically. Any CO introduced into the sample cell will absorb radiation, which reduces the temperature and pressure in the detector cell and displaces a diaphram. This displacement is detected electronically and amplified to provide an output signal.

1.2 This method is applicable to the determination of carbon monoxide in ambient air, and to the analysis of gases under pressure.

2. *Range and Sensitivity.*

2.1 Instruments are available that measure in the range of 0 to 58 mg./m.³ (0–50 p.p.m.), which is the range most commonly used for urban atmospheric sampling. Most instruments measure in additional ranges.

2.2 Sensitivity is 1 percent of full-scale response per 0.6 mg. CO/m.³ (0.5 p.p.m.).

3. *Interferences.*

3.1 Interferences vary between individual instruments. The effect of carbon dioxide interference at normal concentrations is minimal. The primary interference is water vapor, and with no correction may give an interference equivalent to as high as 12 mg. CO/m.³ Water vapor interference can be minimized by (a) passing the air sample through silica gel or similar drying agents, (b) maintaining constant humidity in the sample and calibration gases by refrigeration, (c) saturating the air sample and calibration gases to maintain constant humidity or (d) using narrowband optical filters in combination with some of these measures.

3.2 Hydrocarbons at ambient levels do not ordinarily interfere.

4. *Precision, Accuracy, and Stability.*

4.1 Precision determined with calibration gases is ±0.5 percent full scale in the 0–58 mg./m.³ range.

4.2 Accuracy depends on instrument linearity and the absolute concentrations of the calibration gases. An accuracy of ±1 percent of full scale in the 0–58 mg./m.³ range can be obtained.

4.3 Variations in ambient room temperature can cause changes equivalent to as much as 0.5 mg. CO/m.³ per °C. This effect can be minimized by operating the analyzer in a temperature-controlled room. Pressure changes between span checks will cause changes in instrument response. Zero drift is usually less than ±1 percent of full scale per 24 hours, if cell temperature and pressure are maintained constant.

5. *Apparatus.*

5.1 *Carbon Monoxide Analyzer.* Commercially available instruments should be installed on location and demonstrated, preferably by the manufacturer, to meet or exceed manufacturers specifications and those described in this method.

5.2 *Sample Introduction System.* Pump, flow control valve, and flowmeter.

5.3 *Filter (In-line).* A filter with a porosity of 2 to 10 microns should be used to keep large particles from the sample cell.

5.4 *Moisture Control.* Refrigeration units are available with some commercial instruments for maintaining constant humidity. Drying tubes (with sufficient capacity to operate for 72 hours) containing indicating silica gel can be used. Other techniques that prevent the interference of moisture are satisfactory.

6. *Reagents.*

6.1 *Zero Gas.* Nitrogen or helium containing less than 0.1 mg. CO/m.³

6.2 *Calibration Gases.* Calibration gases corresponding to 10, 20, 40, and 80 percent of full scale are used. Gases must be provided with certification or guaranteed analysis of carbon monoxide content.

6.3 *Span Gas.* The calibration gas corresponding to 80 percent of full scale is used to span the instrument.

7. *Procedure.*

7.1 Calibrate the instrument as described in 8.1. All gases (sample, zero, calibration, and span) must be introduced into the entire analyzer system. Figure C1 shows a typical flow diagram. For specific operating instructions, refer to the manufacturer's manual.

8. *Calibration.*

8.1 *Calibration Curve.* Determine the linearity of the detector response at the operating flow rate and temperature. Prepare a calibration curve and check the curve furnished with the instrument. Introduce zero gas and set the zero control to indicate a recorder reading of zero. Introduce span gas and adjust the span control to indicate the proper value on the recorder scale (e.g. on 0–58 mg./m.³ scale, set the 46 mg./m.³ standard at 80 percent of the recorder chart). Recheck zero and span until adjustments are no longer necessary. Introduce intermediate calibration gases and plot the values obtained. If a smooth curve is not obtained, calibration gases may need replacement.

9. *Calculations.*

9.1 Determine the concentrations directly from the calibration curve. No calculations are necessary.

9.2 Carbon monoxide concentrations in mg./m.³ are converted to p.p.m. as follows:

$$\text{p.p.m. CO} = \text{mg. CO/m.}^3 \times 0.873$$

10. *Bibliography.*

The Intech NDIR–CO Analyzer by Frank McElroy. Presented at the 11th Methods Conference in Air Pollution, University of California, Berkeley, Calif., April 1, 1970.

Jacobs, M. B. et al., J.A.P.C.A. *9, No. 2,* 110–114, August 1959.

MSA LIRA Infrared Gas and Liquid Analyzer Instruction Book, Mine Safety Appliances Co., Pittsburgh, Pa.

Beckman Instruction 1635B, Models 215A, 315A and 415A Infrared Analyzers, Beckman Instrument Company, Fullerton, Calif.

Continuous CO Monitoring System, Model A 5611, Intertech Corp., Princeton, N.J.

Bendix—UNOR Infrared Gas Analyzers. Ronceverte, W. Va.

ADDENDA

A. *Suggested Performance Specifications for NDIR Carbon Monoxide Analyzers:*

Range (minimum)	0–58 mg./m.³ (0–50 p.p.m.).
Output (minimum)	0–10, 100, 1,000, 5,000 mv. full scale.
Minimum detectable sensitivity.	0.6 mg./m.³ (0.5 p.p.m.).
Lag time (maximum)	15 seconds.
Time to 90 percent response (maximum).	30 seconds.
Rise time, 90 percent (maximum).	15 seconds.
Fall time, 90 percent (maximum).	15 seconds.
Zero drift (maximum)	3 percent/week, not to exceed 1 percent/24 hours.
Span drift (maximum)	3 percent/week, not to exceed 1 percent/24 hours.
Precision (minimum)	±0.5 percent.
Operational period (minimum).	3 days.
Noise (maximum)	±0.5 percent.
Interference equivalent (maximum).	1 percent of full scale.
Operating temperature range (minimum).	5–40° C.
Operating humidity range (minimum).	10–100 percent.
Linearity (maximum deviation).	1 percent of full scale.

B. *Suggested Definitions of Performance Specifications:*

Range—The minimum and maximum measurement limits.

Output—Electrical signal which is proportional to the measurement; intended for connection to readout or data processing devices. Usually expressed as millivolts or milliamps full scale at a given impedance.

Full Scale—The maximum measuring limit for a given range.

Minimum Detectable Sensitivity—The smallest amount of input concentration that can be detected as the concentration approaches zero.

Accuracy—The degree of agreement between a measured value and the true value; usually expressed as ± percent of full scale

Lag Time—The time interval from a step change in input concentration at the instrument inlet to the first corresponding change in the instrument output.

Time to 90 percent Response—The time interval from a step change in the input concentration at the instrument inlet to a reading of 90 percent of the ultimate recorded concentration.

Rise Time (90 percent)—The interval between initial response time and time to 90 percent response after a step increase in the inlet concentration.

Fall Time (90 percent)—The interval between initial response time and time to 90 percent response after a step decrease in the inlet concentration.

Zero Drift—The change in instrument output over a stated time period, usually 24 hours, of unadjusted continuous operation, when the input concentration is zero; usually expressed as percent full scale.

Span Drift—The change in instrument output over a stated time period, usually 24 hours, of unadjusted continuous operation, when the input concentration is a stated upscale value; usually expressed as percent full scale.

Precision—The degree of agreement between repeated measurements of the same concentration, expressed as the average deviation of the single results from the mean.

Operational Period—The period of time over which the instrument can be expected to operate unattended within specifications.

Noise—Spontaneous deviations from a mean output not caused by input concentration changes.

Interference—An undesired positive or negative output caused by a substance other than the one being measured.

Interference Equivalent—The portion of indicated input concentration due to the presence of an interferent.

Operating Temperature Range—The range of ambient temperatures over which the instrument will meet all performance specifications.

Operating Humidity Range—The range of ambient relative humidity over which the instrument will meet all performance specifications.

Linearity—The maximum deviation between an actual instrument reading and the reading predicted by a straight line drawn between upper and lower calibration points.

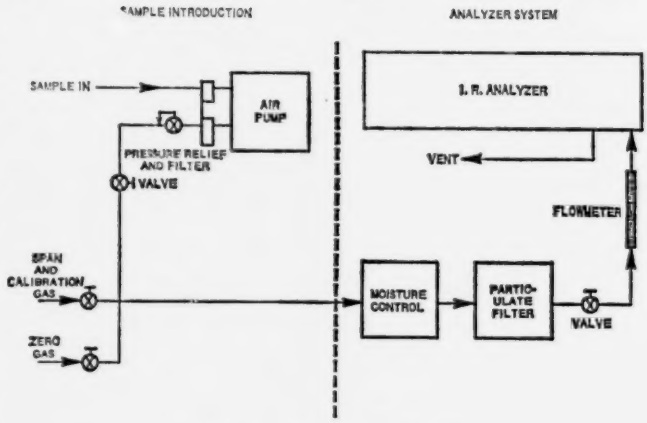

Figure C1. Carbon monoxide analyzer flow diagram.

APPENDIX B
STATISTICAL DESIGN AND ANALYSIS

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	B-1
	A. Purpose and Scope of the Experiment	B-1
	B. Design of the Experiment	B-1
	C. Presentation of the Data	B-2
II.	STATISTICAL ANALYSIS	B-3
	A. Replication Error	B-3
	B. Humidity Effects	B-5
	C. Linear Model Analysis	B-6
III.	INTERPRETATION OF THE PARAMETERS	B-12
	A. Precision of the Method	B-12
	B. Accuracy of the Method	B-15
	LIST OF REFERENCES	B-20

LIST OF ILLUSTRATIONS

Figure		Page
B-1	Design of Carbon Monoxide Method Experiment	B-2
B-2	Control Charts for Means, Slopes, and Standard Errors of Estimate for Linear Model Analysis	B-9
B-3	Repeatability and Reproducibility Versus Concentration	B-12
B-4	Expected Agreement Between Two Means Versus Concentration for Various Numbers of Observations (95 Percent Level of Significance)	B-16
B-5	Expected Agreement Between a Mean and a Fixed Value Versus Concentration for Various Numbers of Observations (95 Percent Level of Significance)	B-17
B-6	Bias or Systematic Error Versus Concentration	B-18

LIST OF TABLES

Table		Page
B-I	Means and Standard Deviations for Each Cell of Data from the Collaborative Test	B-4
B-II	Differences Between Humidified and Dry Test Results	B-5
B-III	Test of Hypothesis That the Mean Difference Between Humidified and Dry Samples Is Equal to Zero	B-6
B-IV	Means and Standard Deviations for Each Laboratory for Each Sample	B-8
B-V	Means, Slopes, and Standard Errors of Estimate for Linear Model Analysis	B-9
B-VI	Analysis of Variance for Linear Model	B-10
B-VII	Summary of Results for Variance Components and Derived Quantities for Linear Model Analysis	B-11
B-VIII	Sources of Variability and Their Relative Importance for the Linear Model Analysis	B-11
B-IX	Standard Errors of Estimate for Calibration Curves Prepared from Calibration Gases and from Reference Gases	B-19

APPENDIX B

STATISTICAL DESIGN AND ANALYSIS

I. INTRODUCTION

In the application of interlaboratory testing techniques, the first step is to determine the exact purpose of the program. There are many, and the particular one must be established. All subsequent details of the program must be planned keeping the prime objective in mind. This appendix describes the design and analysis of the formal collaborative test of the Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry).

A. Purpose and Scope of the Experiment

The basic objective of the interlaboratory study is to derive precise and usable information about the variability of results produced by the measurement method. This information is necessary to establish the reliability of the method in terms of its precision and its accuracy. More emphasis was placed on the inherent quality of the method when properly used than upon the performance of the laboratories.

The statistical planning of the program, which necessarily must be limited in scope, depends upon what information is desired. The scope is limited by what a collaborating laboratory can conveniently and economically accomplish, as well as by the number of collaborators that can be accommodated. Under these limitations, it was possible to examine the effects of laboratories, concentrations, days, and replication upon the precision of the method, in addition to estimating the effects of humidity upon the analysis. The experiment was designed so that the analysis of variance technique could be used.

A total of 16 laboratories took part in the program. An analyst representing each laboratory went through the familiarization phase and subsequently conducted the formal collaborative testing. These individuals and their affiliations have been identified elsewhere in the main report. These laboratories constitute a random sample from a rather large population of experienced laboratories.

Three different concentrations were analyzed by each laboratory. The concentrations were nominally 8, 30, and 53 mg/m^3. These concentrations were selected to approximate the low range, the intermediate range, and the high range for the method. Due to variations among the test gas cylinders, it was not possible for each laboratory to have test atmospheres having the exact values above; however, the expected concentrations are known with confidence, and the deviations of the observed values from the expected values may be examined.

In addition to the analysis of the dry gases, each of the three concentrations was analyzed after humidification according to the technique illustrated in the main report. This information was for the purpose of testing the effectiveness of the various moisture compensation options used in the method.

It was desirable to retrieve information which would allow the investigation of various steps within the method; therefore, emphasis was placed upon obtaining intermediate data relating to calibration curves, moisture compensation methods, instrument ranges, instrument models, and sources of calibration gas. As a result, a substantial amount of data was obtained in addition to the end result of the analytical procedure.

B. Design of the Experiment

A properly planned collaborative test should allow the analysis of the results by the analysis of variance technique or by a procedure which incorporates this technique.[1-5] In general, analysis of variance techniques are more efficient than the simpler control chart techniques. Since the cost of statistical analysis is small compared to the total cost involved in a collaborative test, it is desirable to use the most

efficient statistical methods available in analyzing the results. High efficiency in data utilization is important if the amount of data is limited.

The form of the analysis depends upon the statistical model under consideration. Several separate statistical analyses were performed in order to determine the necessary parameters. Each of these analyses will be described in detail in later subsections.

The overall design of the experiment can best be shown by the diagram in Figure B-1. It can be seen that one analyst in each of 15 laboratories analyzed, in triplicate, each of three concentrations, both dry and humidified, on each of three separate days, resulting in a total of 810 individual determinations. Independent calibration curves were used on each day. The data are presented appropriately in the next subsection. In collaborative testing, two general sources of variability can be readily detected. First, the variability between laboratories can be estimated. This is frequently the largest source of variability and is not under the control of the investigator. Second, the within-laboratory variability can be estimated. This source is under the control of the investigator to the extent that the separate components which make up this source may be identified separately. These separate components, of varying magnitude and importance, may be measured if the proper design has been employed. Alternatively, the separate sources may be confounded or lumped into a single variable by altering the design. By employing the design above, separate estimates could be made of the variability between days and of the variability between replicates. These two components, appropriately combined, constitute the within-laboratory source of variability.

Additional assumptions and rationale for each of the analyses listed previously will be stated later as the analysis is described and applied. If appropriate, the statistical model will be stated in the respective discussion.

C. Presentation of the Data

The data resulting from the experiment are rather voluminous; however, it is essential that these data be tabulated for future reference. In addition to their necessity as supporting information for the problem at hand, the data are also valuable academically as a source of data for the development, evaluation, and comparison of new statistical techniques. Therefore the more voluminous raw data will be found in Appendix C. Data subsets and averages will be presented in this appendix, as appropriate, along with the discussion of the respective statistical analysis.

In presenting the data, all identifiable arithmetic errors have been corrected. The data of Laboratory

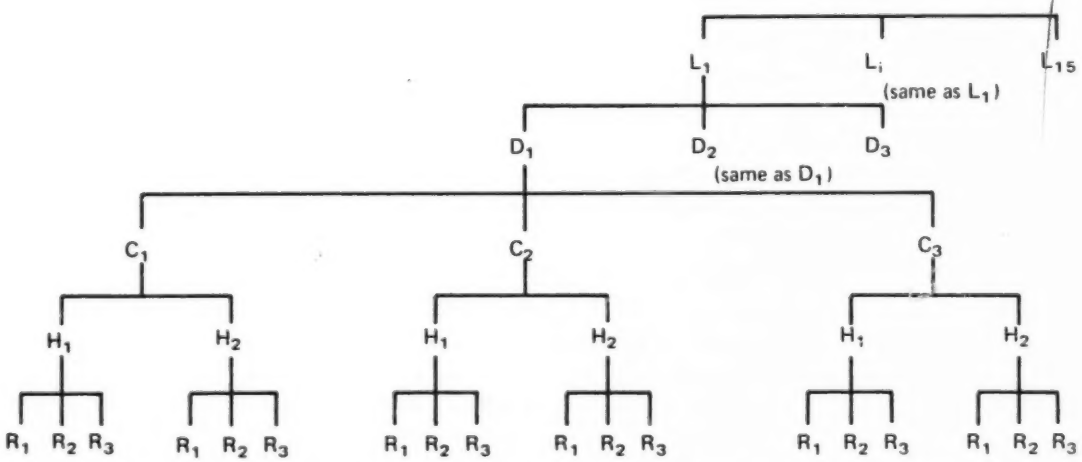

FIGURE B-1. DESIGN OF CARBON MONOXIDE METHOD EXPERIMENT. L, LABORATORIES; D, DAYS; C, CONCENTRATIONS; H, HUMIDITIES; AND R, REPLICATES

780 have been recalculated omitting an extremely suspicious calibration gas. The raw data from 15 of the 16 collaborating laboratories can be seen in Table C-I. One laboratory failed to report complete and usable results.

It is not convenient to have a separate subsection concerning the tests for and disposition of outlying observations. Since there were several statistical analyses and several types of outliers, the outlying observations, if any, will be identified in the respective analysis.

II. STATISTICAL ANALYSIS

Because this study incorporates several statistical techniques, each with its own common notation, certain complications involving consistency of mathematical notation arise. To minimize the confusion, a foldout notation guide is provided at the end of this report. The symbols have been categorized according to their respective use; however, duplicate entries were avoided where there was no conflict or inconsistency. This guide will materially assist the reader throughout this report and may be left folded out for ready reference at any time.

A. Replication Error

To avoid ambiguity, the term *replication* must be explicitly defined. In the context of this study, replicates are defined to be successive determinations with the same operator and instrument on the same sample within intervals short enough to avoid change of environmental factors, and with no intervening manipulations other than zero adjustment. In general, this will mean that time intervals between successive replicates will be on the order of a few to several minutes. Defined in this way, the replication error will primarily reflect the effects of instrument characteristics such as sensitivity, response time, and readout noise. The effects of changing environmental factors will not be included.

The determination of the replication error is straightforward from the data in Table C-1. Each cell provides its own estimate of the replication error with two degrees of freedom. The desired replication error is the combined estimate of these individual estimates, but the question is whether and how these should be combined. Three factors must be investigated in order to answer this question. First, outlying observations must be identified and dealt with. Second, it must be determined if the replication error is a function of concentration. Third, it must be determined if the replication error is affected by humidity. The techniques for each of these analyses are discussed in the following paragraphs.

The means and standard deviations for each cell are shown in Table B-I. The data from laboratories not running replicates consistent with the previous definition are marked with an asterisk and cannot be used in the estimation of the replication error. The data from laboratories not reporting results to the nearest tenth of a milligram per cubic meter were also omitted and are marked with a dagger. It would not be consistent to estimate standard deviations of the magnitudes involved from data rounded to the nearest unit. Inspection of the remaining standard deviations in Table B-I reveal several suspicious results. These remaining data were tested for outliers by Cochran's test[6] applied to each column of standard deviations in Table B-I. The observations thus identified as outliers (99 percent level of significance) have been marked with a double dagger in the table. The values at the foot of each column of standard deviations indicate the magnitude of the pooled estimate and its degrees of freedom for the respective column. The pooled estimates were computed according to usual practice.[7] These results indicate that the replication error is not affected by concentration or by humidity within the limits included in the test. Statistical tests and regression analysis, although hardly necessary, verify this conclusion. Therefore, all individual estimates may be pooled into the final estimate for the replication error σ_ϵ which is 0.17 mg/m^3 (286 degrees of freedom).

Beyond this point, the individual values within each cell are no longer required and all further analyses of the data are made using only the cell averages.

TABLE B-I. MEANS AND STANDARD DEVIATIONS FOR EACH CELL OF DATA FROM THE COLLABORATIVE TEST. First figure in each cell is the mean and the second figure is the standard deviation.

Laboratory Code Number	Day	Low Concentration				Intermediate Concentration				High Concentration			
		Dry		Humid		Dry		Humid		Dry		Humid	
220	1	8.4	0.00	8.6	0.00	30.5	0.06	30.4	0.00	53.2	0.23	53.4	0.06
	2	8.5	0.12	8.6	0.06	30.3	0.21	30.2	0.15	52.3	0.35	52.3	0.35
	3	8.6	0.00	8.7	0.06†	30.5	0.00	30.5	0.00	53.3	0.00	52.8	0.12
222	1	6.9	0.00†	7.0	0.64*	28.6	0.00†	28.6	0.45*	52.7	0.00†	52.7	0.00*
	2	6.9	0.00*	7.4	0.00*	29.0	0.35*	29.8	0.60*	53.1	0.35*	54.0	0.87*
	3	6.9	0.00†	7.2	0.29†	29.0	0.69†	28.1	1.53†	53.5	0.86†	53.3	0.00†
253	1	8.0	0.00†	8.0	0.00†	31.1	0.35†	30.9	0.00†	53.8	0.00†	53.8	0.00†
	2	9.2	0.00†	9.0	0.35†	32.1	0.00†	32.1	0.00†	53.8	0.00†	54.8	0.35†
	3	8.0	0.00†	8.0	0.00†	30.9	0.00†	30.9	0.00†	53.8	0.00†	54.2	0.35†
270	1	7.3	0.64*	7.7	1.33*	30.5	0.64*	31.7	1.33*	56.5	2.66*	56.1	1.96*
	2	8.0	1.15*	8.0	0.00*	31.3	0.69*	32.1	0.00*	55.4	0.64*	55.4	0.64*
	3	8.8	0.69*	8.4	0.69*	31.7	0.69*	31.7	0.69*	55.0	0.00*	55.0	0.00*
310	1	9.0	0.35‡	8.3	0.31	31.3	0.17	30.3	0.23	55.6	0.00	53.9	0.17
	2	9.1	0.17	8.8	0.12	31.2	0.30	30.6	0.25	55.3	0.31	54.5	0.12
	3	9.1	0.31	8.8	0.00	31.3	0.17	31.4	0.17	55.9	0.17	55.3	0.23
311	1	9.1	0.21*	9.5	0.15*	31.4	0.10*	31.2	0.06*	53.4	0.06*	52.8	0.10*
	2	8.6	0.06*	9.3	0.00*	31.5	0.06*	31.2	0.00*	54.3	0.10*	53.4	0.06*
	3	9.5	0.12*	9.7	0.25*	31.0	0.06*	30.9	0.10*	53.5	0.17*	52.9	0.06*
370	1	8.2	0.00	8.3	0.23	30.6	0.00	30.6	0.06	54.4	0.00	54.4	0.00
	2	8.6	0.00	8.6	0.00	30.6	0.00	30.7	0.15	54.2	0.00	54.2	0.06
	3	8.0	0.00	8.1	0.12	30.0	0.35	29.9	0.17	53.4	0.12	53.4	0.12
375	1	7.3	0.00	7.2	0.06	30.3	0.12	29.6	0.23	54.1	0.00	53.4	0.59‡
	2	7.2	0.06	7.2	0.06	29.6	0.00	29.6	0.06	53.4	0.20	53.3	0.12
	3	7.7	0.10	7.7	0.06	30.8	0.17	30.4	0.17	54.2	0.06	53.2	0.20
540	1	8.6	0.12	8.6	0.06	30.3	0.29	30.0	0.06	53.1	0.17	53.3	0.00
	2	8.9	0.12	8.8	0.20	30.5	0.15	30.5	0.15	52.8	0.42‡	52.8	0.10
	3	8.9	0.16	8.6	0.06	30.5	0.06	30.5	0.06	52.9	0.06	53.0	0.21
571	1	8.2	0.00	7.9	0.23	29.4	0.32	29.2	0.06	49.7	0.12	49.3	0.38
	2	8.6	0.05	8.6	0.06	30.2	0.35	29.9	0.17	50.8	0.17	50.4	0.30
	3	8.9	0.06	8.7	0.40	30.6	0.17	30.3	0.17	51.7	0.35	50.9	0.23
780	1	8.1	0.10	12.0	0.10	30.7	0.21	33.4	0.35	52.6	0.26	54.3	0.00
	2	8.8	0.15	12.3	0.25	30.5	0.15	33.6	0.17	52.7	0.15	54.5	0.00
	3	8.2	0.25	11.9	0.12	30.5	0.23	32.8	0.17	53.2	0.15	54.3	0.31
799	1	7.9	0.29	8.2	0.00	29.9	0.00	29.8	0.85‡	53.2	0.23	52.4	0.31
	2	8.3	0.31	8.0	0.00	29.5	0.31	29.2	0.60‡	54.3	0.12	50.9	0.17
	3	8.0	0.00	9.0	0.64‡	31.1	0.17	31.0	0.17	54.7	0.23	54.4	0.17
860	1	7.4	0.00†	7.4	0.00†	30.4	0.00†	30.4	0.00†	52.7	0.00†	52.7	0.00†
	2	7.4	0.00†	7.4	0.00†	30.4	0.00†	30.4	0.00†	52.7	0.00†	52.7	0.00†
	3	7.4	0.00†	7.4	0.00†	30.4	0.00†	30.4	0.00†	52.7	0.00†	52.7	0.00†
920	1	8.0	0.00†	8.0	0.00†	30.2	0.64†	30.5	0.64†	53.8	0.00†	53.8	0.00†
	2	8.0	0.00†	8.0	0.00†	32.1	0.00†	31.7	0.69†	55.0	0.00†	54.6	0.69†
	3	8.0	0.00†	8.0	0.00†	32.1	0.00†	32.1	0.00†	55.0	0.00†	53.8	0.00†
927	1	8.9	0.00	8.9	0.00	32.1	0.00†	32.1	0.00†	55.0	0.00†	55.0	0.00†
	2	8.9	0.00	8.9	0.00	32.1	0.00†	32.1	0.00†	55.0	0.00†	55.0	0.00†
	3	8.9	0.00	8.9	0.00	32.1	0.00†	32.1	0.00†	55.0	0.00†	55.0	0.00†
Mean		0.13		0.20		0.19		0.15		0.17		0.20	
D.F.		26		24		23		25		22		23	

*Not consecutive replicates.
†Results not reported to nearest 0.1 mg/m³.
‡Outlying observations.

B. Humidity Effects

Since humidity is a known interference, the experiment was designed to measure the effectiveness of the various methods of humidity compensation listed in the method (see Section 3.1 of the method in Appendix A). The result of the design was to provide two levels for this factor—one dry and the other essentially saturated. The technique for this humidification step has been discussed in the main report.

Three of the four options for humidity compensation listed in the method were used in the collaborative test. No potential collaborator reported the use of option (c) which is "saturating the air sample and calibration gases to maintain constant humidity." Therefore, this option could not be included. Five laboratories used option (a) using drying agents, and six laboratories used option (b) using refrigeration. Four laboratories used option (d) using narrow-band optical filters. Two of these laboratories used optical filters alone, and two used optical filters in combination with other methods. These laboratories will be identified subsequently when the data are presented.

The effects of humidity can best be determined by pairing the data within days and within concentrations since they are not independent pairs. Statistical techniques to analyze these differences test whether the mean difference is significantly different from zero.[8] In addition to analysis of all differences together, the three subclasses of humidity compensation methods may be analyzed separately to determine whether there are differences in the effectiveness of the respective humidity compensation methods.

The data are shown in Table B-II where the entries are the differences in the means of the three replicates for each humidity level. Each entry is identified according to its respective humidity compensation method as shown by the symbols and their respective footnotes. These data can be shown to be not normally distributed—either overall or within concentrations. The data from two laboratories, 780 and 799, make the major contribution to non-normality. These two laboratories used optical filters and it is obvious that they are not completely effective; however, the one large negative departure for Laboratory 799 is probably an outlier. The data for laboratories using optical filters are not further analyzed due to the small number of laboratories using this method; however, it appears that the use of optical filters in combination with other methods gives satisfactory results.

The data for options (a) and (b), which are normally distributed, are analyzed separately and the

TABLE B-II. DIFFERENCES BETWEEN HUMIDIFIED AND DRY TEST RESULTS. The figures for each day for each concentration for each laboratory are the result of subtracting the dry result from the humidified result, each the average of triplicates. Differences are in milligrams per cubic meter.

Laboratory Code Number	Low Concentration			Intermediate Concentration			High Concentration		
220[b]	0.2	0.1	0.1	−0.1	−0.1	0.0	0.2	0.0	−0.5
222[b]	0.1	0.5	0.3	0.0	0.8	−0.9	0.0	0.9	−0.2
253[a]	0.0	−0.2	0.0	−0.2	0.0	0.0	0.0	1.0	0.4
270[b]	0.4	0.0	−0.4	1.2	0.8	0.0	−0.4	0.0	0.0
310[b]	−0.7	−0.3	−0.3	−1.0	−0.6	0.1	−1.7	−0.8	−0.6
311[d]	0.4	0.7	0.2	−0.2	−0.3	−0.1	−0.6	−0.9	−0.6
370[b]	0.1	0.0	0.1	0.0	0.1	−0.1	0.0	0.0	0.0
375[a,d]	−0.1	0.0	0.0	−0.7	0.0	−0.4	−0.7	−0.1	−1.0
540[b]	0.0	−0.1	−0.3	−0.3	0.0	0.0	0.2	0.0	0.1
571[a]	−0.3	0.0	−0.2	−0.2	−0.3	−0.3	−0.4	−0.4	−0.8
780[d]	3.9	3.5	3.7	2.7	3.1	2.3	1.7	1.8	1.1
799[a,d]	0.3	−0.3	1.0	−0.1	−0.3	−0.1	−0.8	−3.4	−0.3
860[a]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
920[a]	0.0	0.0	0.0	0.3	−0.4	0.0	0.0	−0.4	−1.2
927[a]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[a]Passing the air sample through silica gel or similar drying agent.
[b]Maintaining constant humidity in the sample and calibration gasses by refrigeration.
[d]Using narrow-band optical filters in combination with other measures.

TABLE B-III. TEST OF HYPOTHESIS THAT THE MEAN DIFFERENCE BETWEEN HUMIDIFIED AND DRY SAMPLES IS EQUAL TO ZERO

Statistic	Drying Agents				Refrigeration			
	Concentration*				Concentration*			
	A	B	C	All	A	B	C	All
Number of Observations	15	15	15	45	18	18	18	54
Mean Difference	−0.05	−0.07	−0.12	−0.08	−0.01	−0.01	−0.16	−0.06
Standard Deviation	0.10	0.18	0.50	0.31	0.30	0.54	0.53	0.47
t-value	−1.82	−1.62	−0.93	−1.75	−0.16	−0.04	−1.24	−0.90
Degrees of Freedom	14	14	14	44	17	17	17	53

*A is low concentration, B is intermediate concentration, C is high concentration, and All is all concentrations combined.

results are summarized in Table B-III. No values of the t-statistics[8] are significant at the 95 percent level of significance; therefore, the hypothesis of mean differences equal to zero is accepted. Both methods of moisture compensation appear to be equally satisfactory in comparison with the precision capabilities of the method.

An analysis of variance of the differences in Table B-II (omitting Laboratories 780 and 799) indicates a significant variation between laboratories with respect to the variation between days. Both the variation between laboratories and the variation between days appear to be dependent upon concentration; however, the data are erratic in this respect and the results are inconclusive.

Using the previously determined replication error and the preliminary estimates of the precision between days (0.3, 0.4, and 0.5 mg/m^3 for the low, intermediate, and high concentrations, respectively), an examination of the significance of the magnitude of individual differences can be made. According to these estimates, differences of less than 0.9, 1.2, and 1.4 mg/m^3 for the low, intermediate, and high concentrations, respectively, may be accounted for 95 percent of the time by chance alone. Excluding Laboratories 780 and 799, relatively few observations exceed these amounts.

The humidity has no measurable effect upon the accuracy of the method and does not appear to contribute significantly to the precision.

C. Linear Model Analysis

The assumption made in linear model analysis is that systematic differences exist between sets of measurements made by different observers in different laboratories, and that these systematic differences are linear functions of the magnitude of the measurements. Hence, the technique is called "the linear model."[1,3-5] The linear model leads to a simple design, but requires a special method of statistical analysis, geared to the practical objectives of collaborative tests.

The general design is as follows: to each of p laboratories, q materials have been sent for test, and each laboratory has analyzed each material n times. Now, the n determinations made by the *ith* laboratory on the *jth* material constitute what will be denoted as the "*i,j* cell." The n replicates of any particular cell are viewed as a random sample from a theoretically infinite population of measurements within that cell. The laboratories, however, are not considered as a random sample from a larger population of laboratories, but are considered as fixed variables. Therefore, the inferences involving the variability among laboratories is limited, at least theoretically, to

those laboratories participating in the test. The set of values which corresponds to the q materials is viewed as a fixed variable, but each material is considered to be a random selection from a population of materials with the same "value." This model allows for nonconstant, nonrandom differences between laboratories. The method is not as sensitive to outliers as is the conventional analysis of variance where even a single outlier may result in an unusually large interaction term.

This collaborative test has a nested design in order to allow the differentiation between the reproducibility of results made almost simultaneously and that of results obtained on different days. The term *replicate* in the paragraph above includes both the replication error as it has been previously defined in this appendix as well as the within-laboratory between-days precision yet to be determined.

In view of the results of the analysis of humidity effects discussed in the previous subsection, it is appropriate to combine the data from both dry and humidified test concentrations in this linear model analysis. Since humidity has no apparent effect on either precision or accuracy, there is no reason not to combine the data.

The data in Table B-IV provide the basis for the determination of the between-days precision as well as for the subsequent linear model analysis. The means and standard deviations have been computed as follows:

$$\bar{y}_{ij} = \frac{\sum_{k=1}^{k=w} \bar{y}_{ijk}}{w} - c_{ij} + \bar{c}_j \qquad \text{(B-1)}$$

$$s_{ij}^2 = \frac{\sum_{k=1}^{k=w} (\bar{y}_{ijk} - \bar{\bar{y}}_{ij})^2}{w - 1} \qquad \text{(B-2)}$$

where

$$\bar{y}_{ijk} = \frac{\sum_{m=1}^{m=n} y_{ijkm}}{n} \qquad \text{(B-3)}$$

Each $\bar{y}_{ijk}$ is rounded to 0.1 mg/m^3 before subsequent use in Equation (B-1) or (B-2). For the particular model, there are p = 15 laboratories, q = 6 concentrations or samples, w = 3 days, and n = 3 replicates. The values of c_{ij} are shown in Table C-II of Appendix C and the values of $\bar{c}_j$ are 8, 30, and 53 mg/m^3 for the low, intermediate, and high concentrations, respectively. The reference values are the same for both dry and humidified samples; hence there are only three values whereas q is equal to 6.

The foregoing treatment is necessary in order to remove the variations due to the differences in individual test gas concentrations within a given concentration level. In estimating the replication error or the between-days precision, such treatment was not necessary; however, it was required for subsequent analysis involving variations between laboratories.

The data in Table B-IV are arranged by column for samples and by rows for laboratories with two entries for each cell—the upper is the mean and the lower is the standard deviation. Inspection of the standard deviations reveals some suspiciously high values which must be tested to determine whether they are outliers. Each standard deviation has $w - 1$ degrees of freedom, and each column may be examined by Cochran's test.[6] One observation thus identified as an outlier (99 percent level of significance) has been marked with an asterisk. The values at the foot of each column show the value for the pooled estimate for the column and also the respective degrees of freedom computed in accordance with usual practice.[7]

Further examination of the pooled estimates for each column by regression analysis indicates that there is no significant correlation of standard deviation with concentration. Therefore, all individual estimates may be pooled into a single value equal to 0.45 mg/m^3 (178 degrees of freedom).

The value of 0.45 mg/m^3 corresponds to $V(\epsilon)$, which is the replication variance in the context of the general linear analysis model.[3-5] It can be partitioned into the between-days precision and the

TABLE B-IV. MEANS AND STANDARD DEVIATIONS FOR EACH LABORATORY FOR EACH SAMPLE. Upper number in each cell is mean and lower number is standard deviation. Values in milligrams per cubic meter.

Laboratory Code Number	Dry			Humidified		
	Low Concentration	Medium Concentration	High Concentration	Low Concentration	Medium Concentration	High Concentration
220	8.1 0.10	30.5 0.12	53.3 0.55	8.2 0.06	30.5 0.15	53.2 0.55
222	6.3 0.00	28.7 0.23	54.0 0.40	6.6 0.20	28.6 0.87	54.2 0.65
253	7.9 0.69	31.3 0.64	54.5 0.00	7.8 0.58	31.2 0.69	55.0 0.50
270	7.6 0.75	31.3 0.61	56.4 0.78	7.6 0.35	31.9 0.23	56.3 0.56
310	8.6 0.06	31.5 0.06	56.4 0.30	8.1 0.29	31.0 0.57	55.4 0.70
311	8.5 0.45	31.5 0.26	54.1 0.49	8.9 0.20	31.3 0.17	53.4 0.32
370	7.7 0.31	30.5 0.35	54.9 0.53	7.7 0.25	30.5 0.44	54.9 0.53
375	7.2 0.26	30.0 0.60	54.6 0.44	7.2 0.29	29.7 0.46	54.0 0.10
540	8.3 0.17	30.5 0.12	53.6 0.15	8.2 0.12	30.4 0.29	53.7 0.25
571	8.1 0.35	30.1 0.61	51.4 1.00	7.9 0.44	29.8 0.56	50.9 0.82
780	8.0 0.38	30.8 0.12	53.6 0.32	11.7 0.21	33.5 0.42	55.2 0.12
799	7.7 0.21	30.0 0.83	54.8 0.78	8.0 0.53	29.8 0.92	53.3 1.76*
860	7.2 0.00	30.5 0.00	53.4 0.00	7.2 0.00	30.5 0.00	53.4 0.00
920	7.4 0.00	31.7 1.10	55.3 0.69	7.4 0.00	31.6 0.83	54.8 0.46
927	8.5 0.00	32.0 0.00	55.7 0.00	8.5 0.00	32.0 0.00	55.7 0.00
Pooled Estimate D.F.	0.34 30	0.50 30	0.52 30	0.30 30	0.53 30	0.47 28

*Outlying observation.

replication error as defined in this study according to the relationship

$$V(\epsilon) = \sigma_D^2 + \frac{\sigma_\epsilon^2}{n} \qquad \text{(B-4)}$$

To reduce confusion as much as possible, $V(\epsilon)$ will be used to denote the variance for replication in the context of linear model analysis, and σ_ϵ^2 will be used to denote the variance for replication as defined in this study. Solving Equation (B-4) with $V(\epsilon) = 0.45$, $\sigma_\epsilon = 0.17$, and $n = 3$ yields $\sigma_D = 0.44$ mg/m^3 (161 degrees of freedom) for the value of the standard deviation for between-days precision. The effects of environmental factors and calibration procedures are included in this error term.

Since there was no significant correlation of between-days precision with concentration, there was no

need to make any transformation of scale, and the following linear model analysis was thus made upon the means in Table B-IV.

On the assumption of linear relationships among the p laboratories, it follows that the values obtained by each laboratory are linearly related to the corresponding average values of all laboratories. Each of the means in Table B-IV may be plotted versus its respective column mean. This should be a linear function, and the points corresponding to each line may be represented by three parameters: a mean; a slope; and a quantity related to the deviation from linearity, the standard error of estimate. These parameters are determined by a least-squares regression analysis, and the results are shown in Table B-V. The data of Laboratory 780 have been eliminated from this analysis because of the large differences between dry and humidified samples. The linear model analysis by itself will reveal any other laboratory outliers.

TABLE B-V. MEANS, SLOPES, AND STANDARD ERRORS OF ESTIMATE FOR LINEAR MODEL ANALYSIS. (Omitting Laboratory 780) Data in milligrams per cubic meter.

Laboratory Code Number	Mean	Slope	Standard Error of Estimate
220	30.63	0.9697	0.13
222	29.73	1.0248	0.74
253	31.28	1.0083	0.34
270	31.85	1.0482	0.26
310	31.83	1.0226	0.44
311	31.28	0.9686	0.37
370	31.03	1.0150	0.27
375	30.45	1.0129	0.32
540	30.78	0.9762	0.16
571	29.70	0.9277	0.44
799	30.60	0.9936	0.59
860	30.37	0.9932	0.35
920	31.37	1.0244	0.47
927	32.07	1.0148	0.20
Mean	30.93	1.0000	0.41*

*Pooled estimate.

A plot of the lines represented by the means and slopes from Table B-V would result in a relatively tight bundle of straight lines, each line representing a particular laboratory. Only the lines for Laboratories 222 and 571 depart from the cluster enough to be recognized; therefore, the plot was not reproduced in this report. Both the means and the slopes approximate normal distributions, and no outliers can be detected in either.

Inspection of the standard errors of estimate from Table B-V reveals one suspiciously high value; however, these standard errors of estimate have an approximate chi-square distribution and no outliers can be identified.

These data may be more easily compared from the graphic presentation in Figure B-2 where they have been sorted into an ascending order relative to the means. This sorting often reveals effects not readily visible otherwise. Control limits, based upon deviation from linearity, are shown for the means and the slopes. These 95 percent control limits indicate several points to be "out of control." This indicates that the differences between laboratories cannot be

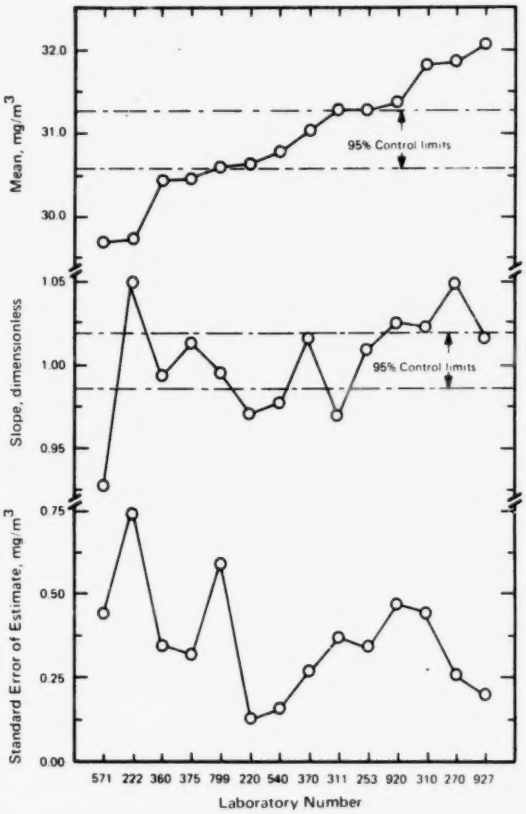

FIGURE B-2. CONTROL CHARTS FOR MEANS, SLOPES, AND STANDARD ERRORS OF ESTIMATE FOR LINEAR MODEL ANALYSIS. (Omitting Laboratory 780).

accounted for by experimental error alone. Examination of Figure B-2 reveals which laboratories showed the greatest departures from the overall mean, which laboratories showed the greatest departure from unit slope, and which laboratories were responsible for the greatest deviations in linearity. When viewing this figure, it is important to watch for relationships between the parameters.

The next step is an analysis of variance which was performed according to the technique of Mandel,[3] and the results are shown in Table B-VI. The interested reader may consult the appropriate reference for the theory and details of the analysis.

The next step in linear model analysis is to determine whether a correlation exists between the means and the slopes. Such a correlation, if it exists, is a valuable feature in the interpretation of the data. The correlation between these two parameters is significant at 90 percent but not at the 95 percent level of significance; therefore, an approximate correlation exists, and the slopes and the means are not completely independent. This significantly positive correlation indicates a tendency for *concurrence* of the lines at a point below the overall mean of 30.9 mg/m^3. If the lines were exactly concurrent, there would exist a particular value of concentration—the point of concurrence—at which all laboratories obtained the same result. An F ratio of the mean square for nonconcurrence to $V(\eta)$ from Table B-VI is highly significant; therefore, the concurrence is not absolute, and there remains a significant amount of variability between laboratories even at the point at which all laboratories tend to agree best. This point lies in the vicinity of zero.

The variance components may now be computed from the data in Table B-VI, and again the technique of Mandel[3] was used. A summary of results for variance components and derived quantities is shown in Table B-VII.

It is now necessary to introduce and define the concept of a *test result*.[9] A *test result* is defined as the average of m replicates, where m is the required number of replicate measurements specified by the method. The particular method does not specify any more than one replicate; therefore the value of m is taken to be one. Thus, a test result is defined as a single measurement and $V(\epsilon)$ is given by Equation (B-4) with $n = m = 1$.

The four sources of variability have been calculated for several values of concentration and are shown in Table B-VIII. Also shown are the fractions of the total variance accounted for by each source. Comparison of $V(\epsilon)$ and $V(\lambda)$, each of which is constant, would indicate that the precision of the method could be improved by decreasing $V(\epsilon)$. However, $V(\epsilon)$ is largely composed of the variation between days, which is large in comparison with the replication error; therefore, increasing the number of replicates will not materially assist in improving the precision of the method. The between-laboratory variability is larger than the within-laboratory variability throughout the table, which indicates significant sources of variation between the laboratories. These sources of variation are undoubtedly related to the accuracy of the calibration gases used in the collaborating laboratories.

The repeatability and reproducibility[9] must now be defined and computed. The repeatability is "a

TABLE B-VI. ANALYSIS OF VARIANCE FOR LINEAR MODEL

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Laboratories	42.6987	13	3.2845
Concentrations	30284.1677	5	6056.8335
Laboratory x Concentration	35.9206	65	0.5526
Linear	27.0714	13	2.0824
Concurrence	7.4996	1	7.4996
Nonconcurrence	19.5718	12	1.6310
Deviation from Linear	8.8491	52	0.1702

TABLE B-VII. SUMMARY OF RESULTS FOR VARIANCE COMPONENTS AND DERIVED QUANTITIES FOR LINEAR MODEL ANALYSIS. Data in milligrams per cubic meter.

Components	Derived from Collaborative Test, $n = w = 3$	For Computations Based on a Test Result, $n = w = 1$
Within Laboratories		
σ_ϵ^2	0.0289	0.0289
σ_D^2	0.1936	0.1936
$V(\epsilon) = \sigma_D^2 + \sigma_\epsilon^2/n$	0.2025	0.2225
$V(\lambda)$	0.1027	0.1027
$V(\eta) = V(\lambda) + V(\epsilon)/w$	0.1702	0.3252
Between Laboratories		
$V(\mu)$	0.5191	0.5191
$V(\beta)$	0.000884	0.000884
$V(\delta)$	0.000754	0.000754
α	0.02207	0.02207
$\bar{x}$	30.9	30.9

quantity that will be exceeded only about five percent of the time by the difference, taken in absolute value, of two randomly selected test results obtained in the same laboratory on a given material."[9] The reproducibility is "a quantity that will be exceeded only about five percent of the time by the difference, taken in absolute value, of two single results made on the same material in two different, randomly selected laboratories."[9] These parameters are computed by the formulas

$$\text{Repeatability} = 2.77\sqrt{V(\eta)} \qquad \text{(B-5)}$$

$$\text{Reproducibility} = 2.77\sqrt{V_j(y)} \qquad \text{(B-6)}$$

where $V(\eta)$ is given by

$$V(\eta) = V(\lambda) + V(\epsilon) \qquad \text{(B-7)}$$

where $V(\epsilon)$ is given by Equation (B-4) for $n = m = 1$ and $V_j(y)$ is given by

$$V_j(y) = \underbrace{(1 + \alpha\gamma_j)^2\ V(\mu) + \gamma_j^2\, V(\delta)}_{\text{Between laboratories}} + \underbrace{V(\eta)}_{\text{Within laboratories}} \qquad \text{(B-8)}$$

where the index j is attached to the variance symbol to signify its dependence upon γ_j which is given by

$$\gamma_j = x_j - \bar{x} \qquad \text{(B-9)}$$

where x_j is the level of concentration at which $V_j(y)$ is desired. Substituting the derived values into Equation (B-8) and simplifying, the following equation is obtained.

$$V_j(y) = 0.001007x_j^2 - 0.0393x_j + 1.10 \qquad \text{(B-10)}$$

Users may choose between Equation (B-8) or (B-10) or the graphic presentation shown in Figure B-3 in which the repeatability and the reproducibility have been plotted for a range of values of concentration.

TABLE B-VIII. SOURCES OF VARIABILITY AND THEIR RELATIVE IMPORTANCE FOR THE LINEAR MODEL ANALYSIS

x	$\sqrt{V(\epsilon)}$	Pct.*	$\sqrt{V(\lambda)}$	Pct.*	$\sqrt{(1+\alpha\gamma)^2 V(\mu)}$	Pct.*	$\sqrt{\gamma^2 V(\delta)}$	Pct.*	$\sqrt{V(y)}$
0	0.45	19	0.32	10	0.23	5	0.85	67	1.04
5	0.45	22	0.32	11	0.31	10	0.71	56	0.95
10	0.45	26	0.32	13	0.39	19	0.57	42	0.89
15	0.45	28	0.32	14	0.47	31	0.44	27	0.85
20	0.45	29	0.32	15	0.55	43	0.30	13	0.83
25	0.45	28	0.32	14	0.63	54	0.16	4	0.85
30	0.45	25	0.32	13	0.71	62	0.03	0	0.90
35	0.45	22	0.32	11	0.79	66	0.11	1	0.97
40	0.45	18	0.32	9	0.86	67	0.25	6	1.06
45	0.45	15	0.32	8	0.94	66	0.39	11	1.16
50	0.45	12	0.32	6	1.02	64	0.52	17	1.28
55	0.45	10	0.32	5	1.10	62	0.66	22	1.40
60	0.45	9	0.32	4	1.18	60	0.80	27	1.53

*Percent of total variance.

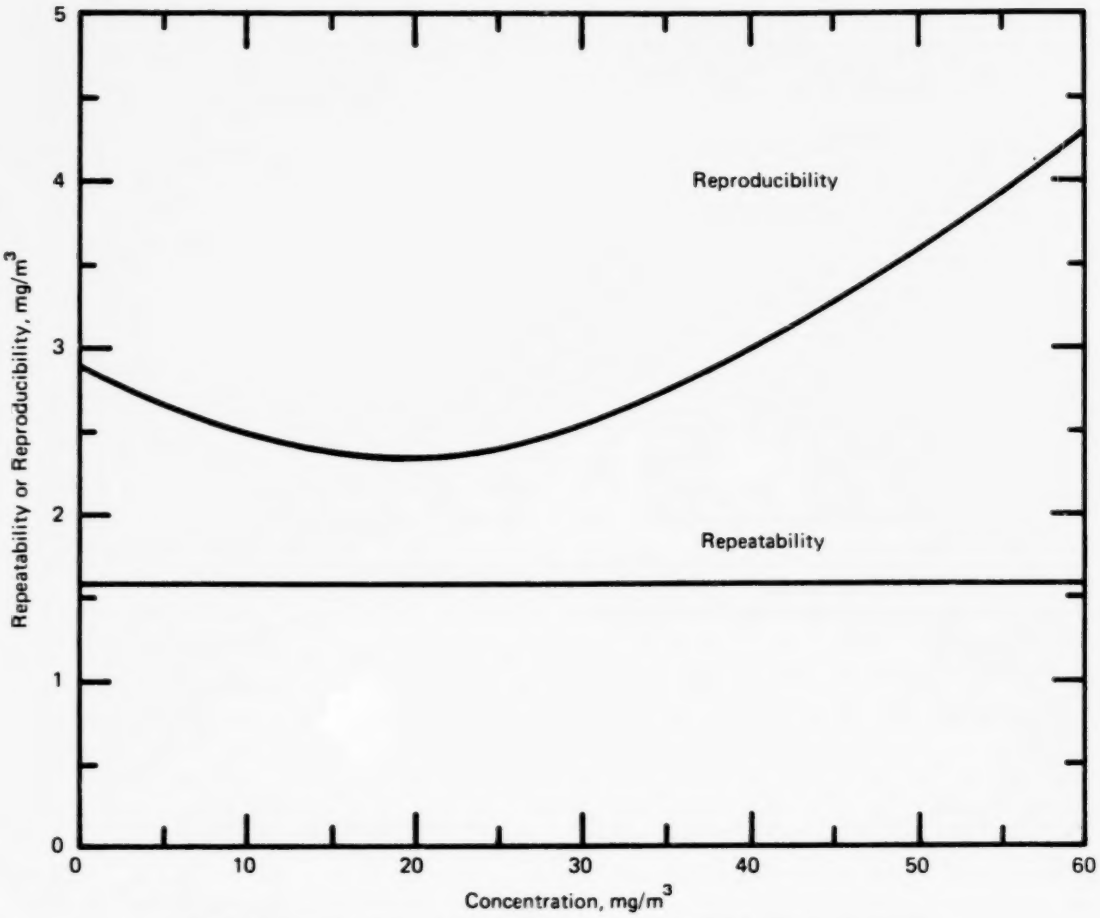

FIGURE B-3. REPEATABILITY AND REPRODUCIBILITY VERSUS CONCENTRATION

III. INTERPRETATION OF THE PARAMETERS

The results of the previous section may now be used to answer some fundamental questions—thus fulfilling the objectives of this collaborative test. Unless otherwise stated below, a 95 percent level of significance is assumed.

A. Precision of the Method

The most general method to test class means is the studentized range.[10-14] If an estimate of the standard deviation s is based on ν degrees of freedom and is independent of the class means to be compared, and if these class means are computed from N cases and selected from a group of g means, then the α allowance for any comparison is

$$\left|\bar{x}_1 - \bar{x}_2\right|_{max} = q_{\alpha(g,\nu)} s/\sqrt{N} \qquad \text{(B-11)}$$

where $\bar{x}_1$ is the highest class mean and $\bar{x}_2$ is the lowest class mean. The value of q is obtained from the appropriate table.[11,13] Interest will center around $g = 2$ because most often the interest is in comparing two class means. In computing checking limits for duplicates, N is of course equal to 1 and the test is identical to ASTM recommended practice.[14]

An obvious limitation is that the means must all contain the same number of observations. When this

is not the case, the standard normal deviate is adequate[15] and use can be made of the equation

$$\left|\bar{x}_1 - \bar{x}_2\right|_{max} = z_{(1 - \frac{1}{2}\alpha)}\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \qquad \text{(B-12)}$$

where $|\bar{x}_1 - \bar{x}_2|$ is the absolute value of the difference in the two class means $\bar{x}_1$ and $\bar{x}_2$, and N_1 and N_2 are the numbers of observations in $\bar{x}_1$ and $\bar{x}_2$, respectively. The results from this equation are the same as Equation (B-11) when $N = N_1 = N_2$ and ν is large. The results are adequate if N_1 and N_2 are relatively large (20 or more).

To test whether the true value of a mean is lower than a specified fixed value, the maximum permissible difference is[16]

$$(\bar{x} - \mu_0)_{max} = z_{(1-\alpha)}\sigma/\sqrt{N} \qquad \text{(B-13)}$$

which is a one-sided test where $\bar{x}$ is the mean, μ_0 is the fixed value, and N is the number of observations in $\bar{x}$.

These techniques will be applied as appropriate to the three sources of variation below. The treatment will be in more depth for the precision between laboratories, which is of more practical interest.

1. Precision Between Replicates

We have already concluded that replication will not materially assist in increasing the precision of the method. Replication will, in general, be a waste of time and effort; however, replicates are often advisable to avoid gross errors. The expression for the checking limit for duplicates uses σ_ϵ and Equation (B-11) yielding

$$R_{max} = 2.77(0.17) = 0.5 \qquad \text{(B-14)}$$

where R_{max} is the maximum permissible range between duplicates. Two such replicates should be considered suspect if they differ by more than 0.5 mg/m^3.

2. Precision Between Days

One situation involves within-laboratory comparisons of the same sample. It is of interest when comparing measured values on the same sample analyzed on separate days. The estimate of the standard deviation in Equation (B-11) must now include the variation between days in addition to the replication error. The expression for R_{max}, the maximum permissible range between two test results, is

$$R_{max} = 2.77\sqrt{V(\epsilon)} = 1.3 \qquad \text{(B-15)}$$

where $V(\epsilon)$ is given by Equation (B-4) for $n = m = 1$. Two such test results should be considered suspect if they disagree by more than 1.3 mg/m^3.

A separate and distinct case arises for within-laboratory comparisons of two samples. Suppose it is desired to compare the results from a single laboratory on two different but similar samples analyzed on different days. The samples may have the same concentration but may differ in other interfering properties such as humidity. It must be assumed that the heterogeneity between the two samples with respect to interfering properties is essentially the same as that shown in the collaborative test. Therefore, the estimate of $V(\lambda)$ is the appropriate measure for the possible heterogeneity of the two samples. Thus, the standard deviation estimate for Equation (B-11) must now include $V(\lambda)$ as well. The resulting expression for R_{max}, the maximum permissible range between the test results on each sample is

$$R_{max} = 2.77\sqrt{V(\lambda) + V(\epsilon)} = 1.6 \qquad \text{(B-16)}$$

Therefore, the maximum permissible difference between a single test result on each of the samples is 1.6 mg/m^3. If two such test results differ by less than 1.6 mg/m^3 there is no reason to believe that there is any real difference between them.

There may also be some occasions where it will be necessary to compare the means for each of two given sampling stations, where each mean was obtained by the same analyst, and consisted of a known number of test results. The number of observations in each mean will not usually be equal. Their standard deviations will not usually be equal, and one or both may not be normally distributed. Where they are normally distributed, standard tests such as the t-test[17] may be applied.

A limiting case may be investigated if it is assumed that two means $\bar{x}_1$ and $\bar{x}_2$ are normally distributed with $\sigma_1 = \sigma_2 = \sqrt{V(\lambda) + V(\epsilon)} = 0.57$ mg/m^3. This is an unlikely, if not impossible, situation which could only result from absolutely constant concentrations at each of the sampling stations. Under these assumptions, we may apply Equation (B-12) and obtain

$$R_{max} = 1.1 \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \qquad \text{(B-17)}$$

where R_{max} is the maximum permissible range between means $\bar{x}_1$ and $\bar{x}_2$ containing N_1 and N_2 observations, respectively. If the range exceeds R_{max}, the means are significantly different and do not belong to the same population.

Under the same limiting assumptions, a mean $\bar{x}$ containing N observations may be compared with some fixed value μ_0 and it may be stated whether the true value of $\bar{x}$ is less than μ_0. Equation (B-13) may be applied to this case resulting in

$$R_{max} = 0.9/\sqrt{N} \qquad \text{(B-18)}$$

where R_{max} is the maximum permissible range between $\bar{x}$ and μ_0. If $\bar{x} - \mu_0$ is less than R_{max}, then the true value of $\bar{x}$ is less than μ_0.

3. Precision Between Laboratories

Probably the most frequent comparison to be made will be that involving observations of two different laboratories. When a comparison is made between results obtained in different laboratories, the variance $V(\lambda)$ is always included in the comparison, regardless of whether this comparison involves a single material or different materials. While it is true that the interfering properties for a single material are constant, the response of different laboratories to the same interfering property may not necessarily be the same. The variability of this response is exactly what is measured by $V(\lambda)$. The estimate of the standard deviation for Equation (B-11) now contains the effects of variations in the means and the slopes of the response lines for the laboratories. The required estimate is the square root of $V_j(y)$ which may be obtained from either Equation (B-8) or (B-10). The resulting expression for R_{max}, the maximum permissible difference between a test result from each of two different laboratories, is

$$R_{max} = 2.77 \sqrt{V_j(y)} \qquad \text{(B-19)}$$

This comparison is complicated by the dependence of between-laboratory variability on the concentration. R_{max} is identical to the reproducibility given by Equation (B-6) and plotted in Figure B-3. Two such test results may not be considered to belong to the same population if they differ by more than R_{max}. Conversely, the two test results are not significantly different if they differ by less than R_{max}.

Frequently, it will be necessary to compare the means for each of two given sampling stations. Each mean may be the result of observations by one or more different laboratories. Each mean may contain a different number of observations, each a test result. Their standard deviations will not usually be equal, and one or both may not be normally distributed. Where they are normally distributed, standard tests such as the t-test[17] may be applied.

Similar to the preceding subsection, a limiting case may be investigated if it is assumed that the two means $\bar{x}_1$ and $\bar{x}_2$ containing $N_1 = N_2 = N$ observations are normally distributed with $\sigma_1 = \sigma_2 = \sqrt{V_j(y)}$. Here again, this is an unlikely, if not impossible, situation which could only result from absolutely constant concentrations at each sampling station. Nevertheless, a certain amount of guidance can be derived. If Equation (B-11) is applied to this case, the result is

$$R_{max} = 2.77 \sqrt{\frac{V_j(y)}{N}} \qquad \text{(B-20)}$$

where R_{max} is the maximum permissible range between the means $\bar{x}_1$ and $\bar{x}_2$. If the range exceeds R_{max}, the means are significantly different and do not belong to the same population.

Under the same assumptions as above, with the exception that N_1 may not equal N_2 but both are relatively large, Equation (B-12) is used, yielding

$$R_{max} = 1.96\sqrt{V_f(y)}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \qquad \text{(B-21)}$$

where R_{max} is the maximum permissible range between $\bar{x}_1$ and $\bar{x}_2$. If the range exceeds R_{max}, the means are significantly different and do not belong to the same population.

It is interesting to pursue this line of reasoning further in terms of the number of samples required to detect a specified difference under the limiting assumptions. Rearranging Equation (B-20) and solving for N, the result is

$$N = \left(\frac{2.77}{R_{max}}\right)^2 V_f(y) \qquad \text{(B-22)}$$

This expression now gives the minimum number of observations N for any desired agreement R_{max} between two means at any level of concentration x_j. These results are best illustrated in Figure B-4. This figure shows the agreement versus the concentration level for a family of sample sizes. Superimposed on the curve are constant percentage agreement lines for comparison purposes. For example, if agreement better than 5 percent at a concentration of 15 mg/m^3 is desired, a minimum of 10 observations would be required.

Under the same limiting assumptions, it is possible to compare a mean $\bar{x}$ containing N observations with some fixed value μ_0 and be able to state whether the true value of $\bar{x}$ is less than μ_0. Equation (B-13) is used for this type case yielding

$$R_{max} = -1.645\sqrt{\frac{V_f(y)}{N}} \qquad \text{(B-23)}$$

where R_{max} is the maximum permissible range between $\bar{x}$ and μ_0. If $\bar{x} - \mu_0$ is less than R_{max}, then the true value of $\bar{x}$ is less than μ_0.

Rearranging Equation (B-23) and solving for N yields

$$N = \left(\frac{1.645}{R_{max}}\right)^2 V_f(y) \qquad \text{(B-24)}$$

This equation is exactly analogous to Equation (B-22). N is the minimum number of observations required to attain the agreement R_{max} under the limiting assumptions. Figure B-5, which is analogous to Figure B-4, best illustrates the resulting relationships. For example, a minimum of two observations would be required to establish that the true value of $\bar{x}$ is less than 20 mg/m^3, while the actual value is 19 mg/m^3 (a 5-percent difference). Stated differently, given a set of two observations with a mean of 19 mg/m^3, there exists a 95 percent confidence that the true mean is less than 20 mg/m^3

B. Accuracy of the Method

In the discussion of accuracy, an additional concept must be introduced–the *reference value* of the measured property for the system under consideration. Mandel[18] discusses three types of reference values of which the "assigned value" applies for this collaborative test. The reference values for the samples included in the study are the values provided for these samples by the supplier of those samples. This does not necessarily mean that these values are considered absolutely correct, but it does mean that there is a reasonable degree of confidence in the quality of such materials from this source.

If the reference value is represented by R and the mean of the population of repeated measurements is μ, then the bias or systematic error is $\mu - R$. The error for an individual measurement x would be $x - R$. Inaccuracy is thus measured by the magnitude of $\mu - R$ or $x - R$. A method is accurate if $\mu - R$ is not significantly different from zero.

A definite and statistically significant inaccuracy exists; however, its practical significance must be interpreted with respect to other criteria. This inaccuracy is best illustrated in Figure B-6. The

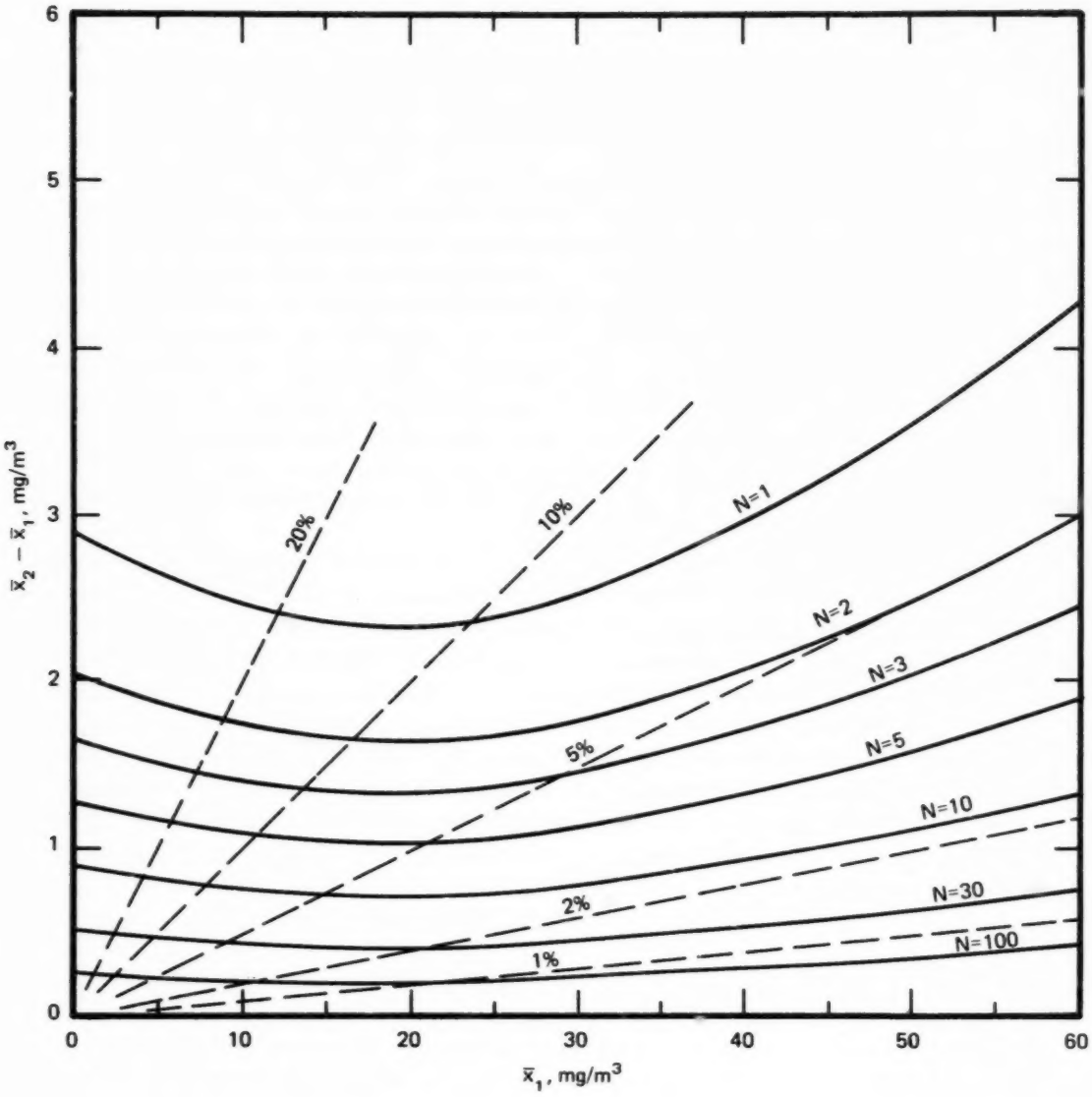

FIGURE B-4. EXPECTED AGREEMENT BETWEEN TWO MEANS VERSUS CONCENTRATION FOR VARIOUS NUMBERS OF OBSERVATIONS (95 Percent Level of Significance). EACH MEAN HAS N OBSERVATIONS WITH A STANDARD DEVIATION EQUAL TO $(0.001007\bar{x}_1^3 - 0.0393\bar{x}_1 + 1.10)^{0.5}$

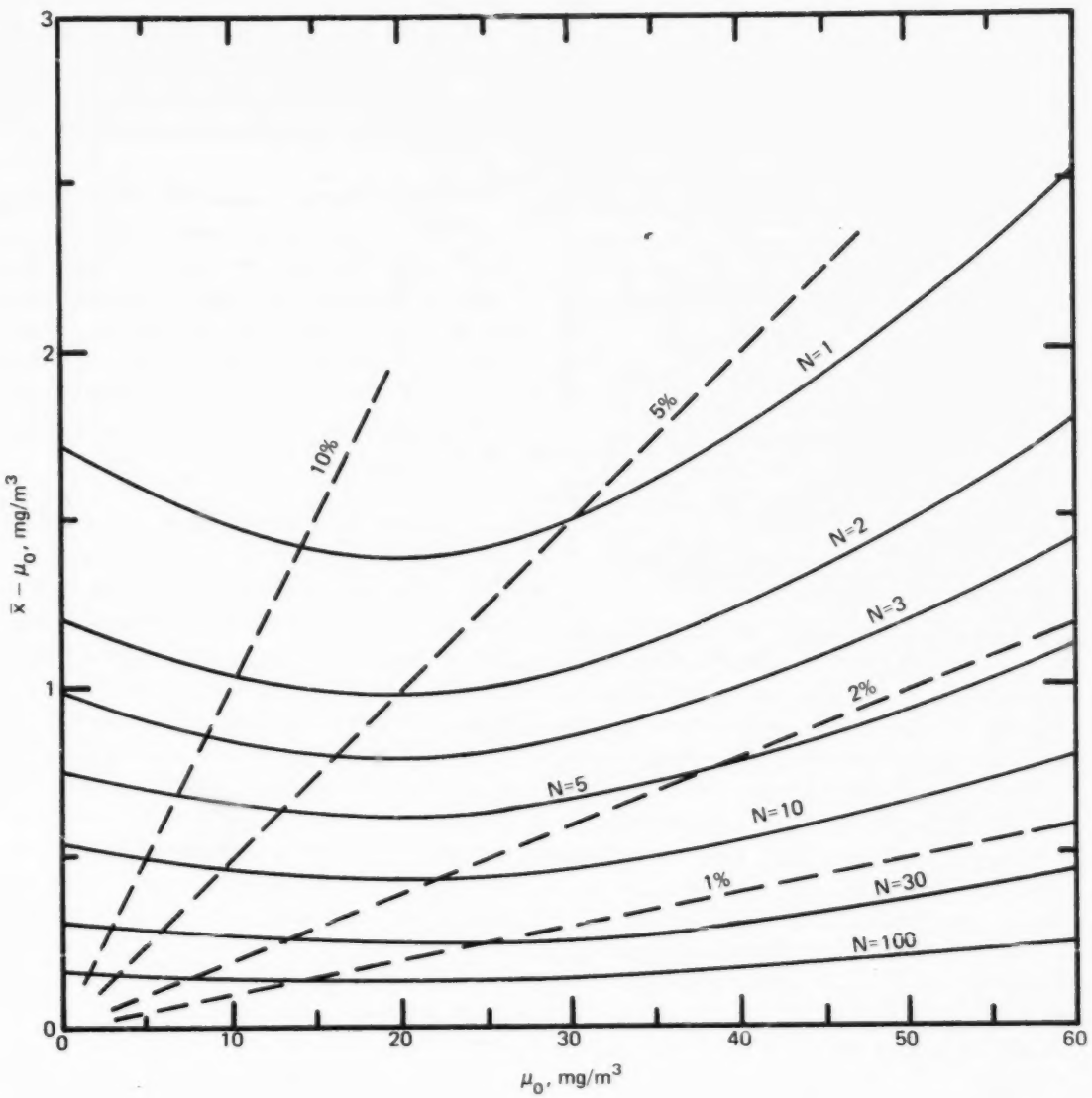

FIGURE B-5. EXPECTED AGREEMENT BETWEEN A MEAN AND A FIXED VALUE VERSUS CONCENTRATION FOR VARIOUS NUMBERS OF OBSERVATIONS (95 Percent Level of Significance). THE MEAN HAS N OBSERVATIONS WITH A STANDARD DEVIATION EQUAL TO $(0.001007\mu_0^2 - 0.0393\mu_0 + 1.10)^{0.5}$

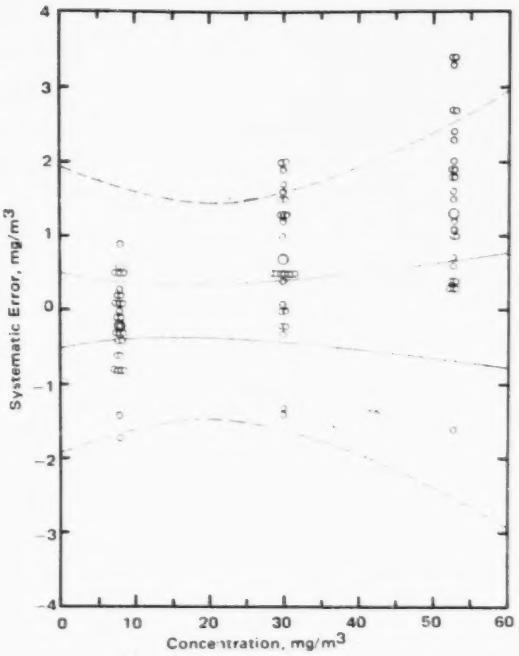

FIGURE B-6. BIAS OR SYSTEMATIC ERROR VERSUS CONCENTRATION

departures of individual laboratory averages from their respective reference values as well as the departures of overall averages from their respective values have been plotted versus concentration. The large open circles are overall averages at each level of concentration in the collaborative test. The small circles are individual laboratory values. The overall averages at each level are the mean of 14 laboratory averages, each of which is the average of 18 observations (three replicates on each of three days for each of two samples). The standard error of the overall means is represented by the solid lines, and the standard error of the individual means is represented by the dashed lines. The standard errors have been plotted with reference to zero so that observations falling outside these lines are significantly different from zero.

Several of the individual laboratory means are significantly different from zero, mostly at the higher concentrations and nearly all on the high side. The overall means are significant at the two higher levels of concentration. This relationship is nearly linear and the results tend to be, on the average, 2.5 percent high.

The method uses the same type materials for calibration as were used for reference samples in this test. There is little doubt, therefore, that the inaccuracy results primarily, if not completely, from the use of calibration gases which exhibit significant variation with respect to their specified concentration. Since the results tend to be high, the calibration gases must have a tendency to be correspondingly low.

Caution should be exercised in the use of these measures of accuracy. Although calibration gas sources were randomly selected, it is known that some standards used by different laboratories were prepared and analyzed at the same time by the same supplier. Nevertheless, it cannot be overemphasized that the accuracy of the method is almost totally dependent upon the availability of sufficiently accurate standards.

In order to further examine the accuracy of the calibration gases used by the individual collaborators, some additional analyses were made. The first of these investigated the individual calibration curves and compared them with calibration curves constructed from the reference sample data. Since the chart readings for each calibration curve were recorded, the parameters of each curve could be computed. The standard error of estimate was computed for each calibration curve by a least-squares regression analysis of chart readings on calibration gas concentration. The average standard errors of estimate for each laboratory are shown in Table B-IX, where they have been grouped into instrument ranges and subgrouped into calibration gas sources. They are in units of chart divisions and, for purposes of comparison, a chart division for the 0 to 58-mg/m^3 (0 to 50 ppm) range is approximately equal to 0.6 mg/m^3, and a chart division for the higher range is approximately equal to 1.2 mg/m^3. These standard errors of estimate are measures of nonlinearity of the calibration curves.

In order to provide individual comparisons, the same analysis was performed on the reference samples and their respective chart readings as though they were actually calibration gases. These are also shown in Table B-IX. Inspection reveals some unusually large

TABLE B-IX. STANDARD ERRORS OF ESTIMATE FOR CALIBRATION CURVES PREPARED FROM CALIBRATION GASES AND FROM REFERENCE GASES

Laboratory Code Number	Instrument*	Range†	Water Vapor Compensation‡	Calibration Gas Source*	Standard Error of Estimate	
					Calibration	Reference
571	A	0-58	a	A	1.0**	1.2**
370	B	0-58	b	B	1.0**	1.1**
375	A	0-58	a & d	B	2.5**	0.7**
799	A	0-58	a & d	C	4.3**	0.9**
310	A	0-58	b	D	1.8**	0.9**
311	A	0-58	d	D	1.2**	1.2**
222	C	0-116	b	E	1.2††	1.9††
220	D	0-116	b	A	0.7††	0.2††
253	B	0-116	a	A	0.4††	0.5††
540	A	0-116	b	A	0.4††	0.3††
860	B	0-116	a	A	0.0††	0.5††
920	A	0-116	a	B	1.6††	0.5††
780	A	0-116	d	C	3.7††	0.4††
927	B	0-116	a	F	1.3††	0.2††
270	A	0-116	b	E	2.4††	0.5††

*Coded to obscure identity.
†Milligrams per cubic meter.
‡See Section 3.1 of method in Appendix A.
**Chart divisions–1 chart division is approximately equivalent to 0.6 mg/m^3.
††Chart divisions–1 chart division is approximately equivalent to 1.2 mg/m^3.

values in the calibration gas data. In the majority of cases, the standard error of estimate for the calibration gases is larger than the corresponding value for the reference samples. It is evident that the calibration gases are more variable than the reference samples; the question remains, to what can this variability be ascribed?

To explore this matter further, each calibration gas was "analyzed," using its respective chart reading and the "calibration curve" prepared from reference samples. Such a treatment corresponds to giving the collaborator the concentrations of the reference samples and asking him to prepare a calibration curve from them and then analyze his own calibration gases as if they were unknown samples. This is most informative since these "analytical results" may be compared with the specified value for the calibration gases. Unusually large differences would point to a suspicious calibration gas.

Following this procedure, some of the large standard errors of estimate can be explained by one suspicious calibration gas. Some notable examples of differences more than 10 percent are (1) Laboratory 799–a higher value than quoted for 23 mg/m^3 (20 ppm) calibration gas, (2) Laboratory 270–a much lower value than quoted for 91 mg/m^3 (80 ppm) calibration gas, (3) Laboratory 780–a lower value than quoted for 46 mg/m^3 (40 ppm) calibration gas, and (4) Laboratory 927–a lower value than quoted for 23 mg/m^3 (20 ppm) calibration gas. Other cases of high standard errors of estimate could not be attributed to a single suspicious calibration gas. Several other cases of differences of 10 percent were observed along with numerous cases of 5-percent differences. Absolute magnitudes of "analyzed" values minus quoted values ranged from -10.2 to $+3.3$ mg/m^3.

It should be noted, however, that some laboratories did not use suspicious calibration points in computing their results for the reference samples and preferred to use their judgment based on the calibration curve as a whole, sometimes drawing nonlinear calibration curves. While this practice can often minimize inaccuracy, it can also lead to worse situations. For instance, Laboratory 780, upon inspection of its calibration curve, thought the 91-mg/m^3 (80 ppm) point to be out of line and chose not to use it when actually the 46-mg/m^3 (40 ppm) calibration gas was at fault.

Utmost care should be taken in obtaining high-quality calibration gases and protecting them from deterioration. If smooth calibration curves are not obtained, calibration gases may be at fault and should be replaced.

LIST OF REFERENCES

1. ASTM Manual for Conducting an Interlaboratory Study of a Test Method, ASTM STP No. 335, *Am. Soc. Testing & Mats.* (1963).

2. **1971 Annual Book of ASTM Standards,** Part 30, Recommended Practice for Developing Precision Data on ASTM Methods for Analysis and Testing of Industrial Chemicals, ASTM Designation: E180-67, pp 403-422.

3. Mandel, John, **The Statistical Analysis of Experimental Data**, John Wiley & Sons, New York, Chapter 13, pp 312-362 (1964).

4. Mandel, J., and Lashof, T.W., "The Interlaboratory Evaluation of Testing Methods," *ASTM Bulletin*, No. 239, pp 53-61 (1959).

5. Mandel, John, "The Measuring Process," **Technometrics**, 1, pp 251-267 (1959).

6. Dixon, Wilfred J., and Massey, Frank J., Jr., **Introduction to Statistical Analysis**, McGraw-Hill Book Company, Inc., New York, Chapter 10, pp 180-181 (1957).

7. Ibid, Chapter 8, pp 109-110.

8. Ibid, Chapter 9, pp 112-129.

9. Mandel, John, "Repeatability and Reproducibility," *Materials Research and Standards, Am. Soc. Testing & Mats.*, Vol 11, No. 8, p 8 (August 1971).

10. Duncan, Acheson J., **Quality Control and Industrial Statistics**, Third Edition, Richard D. Irwin, Inc., Homewood, Illinois, Chapter XXXI, pp 632-636 (1965).

11. Ibid, p 909.

12. Bennett, Carl A., and Franklin, Norman L., **Statistical Analysis in Chemistry and the Chemical Industry**, John Wiley and Sons, Inc., New York, Chapter 4, p 111 (1954).

13. Ibid, p 185-189.

14. **1971 Annual Book of ASTM Standards**, op cit, p 411.

15. Dixon and Massey, op cit, p 120.

16. Ibid, pp 114-115.

17. Ibid, pp 123-124.

18. Mandel, John, **The Statistical Analysis of Experimental Data**, John Wiley & Sons, New York, Chapter 6, pp 104-105 (1964).

APPENDIX C

TABULATION OF ORIGINAL DATA

TABLE C-I-a. OBSERVED VALUES FOR DRY SAMPLES FOR COLLABORATIVE TEST OF CARBON MONOXIDE METHOD, MILLIGRAMS PER CUBIC METER

Laboratory Code Number	Day 1			Day 2			Day 3		
Low Concentration									
220	8.4	8.4	8.4	8.4	8.4	8.6	8.6	8.6	8.6
222	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9
253	8.0	8.0	8.0	9.2	9.2	9.2	8.0	8.0	8.0
270	8.0	6.9	6.9	9.2	8.0	6.9	9.2	9.2	8.0
310	9.4	8.8	8.8	8.9	9.2	9.2	9.4	8.8	9.2
311	8.9	9.3	9.0	8.7	8.6	8.6	9.4	9.6	9.6
370	8.2	8.2	8.2	8.6	8.6	8.6	8.0	8.0	8.0
375	7.3	7.3	7.3	7.2	7.2	7.1	7.6	7.8	7.7
540	8.5	8.5	8.7	9.0	9.0	8.8	8.9	8.9	8.8
571	8.2	8.2	8.2	8.6	8.6	8.7	8.9	8.9	9.0
780	8.2	8.0	8.1	8.9	8.6	8.8	8.0	8.5	8.2
799	7.7	8.2	7.7	8.6	8.2	8.0	8.0	8.0	8.0
860	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.4
920	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
927	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9
Intermediate Concentration									
220	30.5	30.5	30.6	30.1	30.4	30.4	30.5	30.5	30.5
222	28.6	28.6	28.6	29.2	29.2	28.6	28.6	29.8	28.6
253	30.9	30.9	31.5	32.1	32.1	32.1	30.9	30.9	30.9
270	29.8	30.9	30.9	32.1	30.9	30.9	32.1	32.1	30.9
310	31.2	31.2	31.5	31.5	30.9	31.2	31.2	31.2	31.5
311	31.5	31.4	31.3	31.5	31.4	31.5	30.9	31.0	31.0
370	30.6	30.6	30.6	30.6	30.6	30.6	30.4	29.8	29.8
375	30.2	30.2	30.4	29.6	29.6	29.6	30.9	30.9	30.6
540	30.6	30.1	30.1	30.5	30.7	30.4	30.5	30.6	30.5
571	29.3	29.2	29.8	30.4	30.4	29.8	30.7	30.4	30.7
780	30.9	30.6	30.5	30.7	30.5	30.4	30.2	30.6	30.6
799	29.9	29.9	29.9	29.8	29.2	29.4	31.2	30.9	31.2
860	30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4
920	29.8	30.9	29.8	32.1	32.1	32.1	32.1	32.1	32.1
927	32.1	32.1	32.1	32.1	32.1	32.1	32.1	32.1	32.1
High Concentration									
220	53.3	53.3	52.9	52.7	52.1	52.1	53.3	53.3	53.3
222	52.7	52.7	52.7	53.3	53.3	52.7	53.3	52.7	54.4
253	53.8	53.8	53.8	53.8	53.8	53.8	53.8	53.8	53.8
270	59.6	55.0	55.0	56.1	55.0	55.0	55.0	55.0	55.0
310	55.6	55.6	55.6	55.6	55.2	55.0	55.8	56.1	55.8
311	53.4	53.3	53.4	54.3	54.4	54.2	53.6	53.3	53.6
370	54.4	54.4	54.4	54.2	54.2	54.2	53.5	53.3	53.5
375	54.1	54.1	54.1	53.2	53.4	53.6	54.2	54.2	54.3
540	53.3	53.0	53.0	52.5	53.3	52.7	52.9	52.9	52.8
571	49.6	49.6	49.8	51.0	50.7	50.7	51.5	52.1	51.5
780	52.7	52.8	52.3	52.9	52.6	52.7	53.2	53.0	53.3
799	52.9	53.3	53.3	54.4	54.2	54.4	55.0	54.6	54.6
860	52.7	52.7	52.7	52.7	52.7	52.7	52.7	52.7	52.7
920	53.8	53.8	53.8	55.0	55.0	55.0	55.0	55.0	55.0
927	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0

Preceding page blank

TABLE C-I-b. OBSERVED VALUES FOR HUMIDIFIED SAMPLES FOR COLLABORATIVE TEST OF CARBON MONOXIDE METHOD, MILLIGRAMS PER CUBIC METER

Laboratory Code Number	Day 1			Day 2			Day 3		
Low Concentration									
220	8.6	8.6	8.6	8.6	8.7	8.6	8.7	8.7	8.6
222	7.4	6.3	7.4	7.4	7.4	7.4	6.9	7.4	7.4
253	8.0	8.0	8.0	9.2	9.2	8.6	8.0	8.0	8.0
270	9.2	6.9	6.9	8.0	8.0	8.0	8.0	8.0	9.2
310	8.6	8.0	8.2	8.9	8.7	8.7	8.8	8.8	8.8
311	9.7	9.5	9.4	9.3	9.3	9.3	9.7	9.4	9.9
370	8.6	8.2	8.2	8.6	8.6	8.6	8.0	8.0	8.2
375	7.2	7.2	7.3	7.3	7.2	7.2	7.8	7.7	7.7
540	8.6	8.5	8.6	8.6	9.0	8.8	8.6	8.7	8.6
571	7.8	7.8	8.2	8.6	8.6	8.5	8.2	8.9	8.9
780	11.9	12.1	12.0	12.5	12.3	12.0	12.0	11.8	11.8
799	8.2	8.2	8.2	8.0	8.0	8.0	9.7	8.6	8.6
860	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.4
920	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
927	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9
Intermediate Concentration									
220	30.4	30.4	30.4	30.4	30.2	30.1	30.5	30.5	30.5
222	28.6	28.1	29.0	29.2	30.4	29.8	27.5	29.8	26.9
253	30.9	30.9	30.9	32.1	32.1	32.1	30.9	30.9	30.9
270	33.2	30.9	30.9	32.1	32.1	32.1	32.1	32.1	30.9
310	30.0	30.4	30.4	30.6	30.9	30.4	31.5	31.2	31.5
311	31.3	31.2	31.2	31.2	31.2	31.2	30.8	30.9	31.0
370	30.6	30.6	30.7	30.9	30.7	30.6	30.1	29.8	29.8
375	29.3	29.7	29.7	29.7	29.6	29.6	30.5	30.5	30.2
540	30.1	30.0	30.0	30.5	30.4	30.7	30.5	30.4	30.5
571	29.2	29.2	29.3	30.1	29.8	29.8	30.4	30.4	30.1
780	33.7	33.4	33.0	33.7	33.4	33.7	32.9	32.9	32.6
799	30.6	29.9	28.9	29.8	29.2	28.6	30.9	30.9	31.2
860	30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4
920	29.8	30.9	30.9	30.9	32.1	32.1	32.1	32.1	32.1
927	32.1	32.1	32.1	32.1	32.1	32.1	32.1	32.1	32.1
High Concentration									
220	53.4	53.4	53.5	52.7	52.1	52.1	52.9	52.7	52.7
222	52.7	52.7	52.7	53.3	53.8	55.0	53.3	53.3	53.3
253	53.8	53.8	53.8	55.0	54.4	55.0	53.8	54.4	54.4
270	58.4	55.0	55.0	55.0	55.0	56.1	55.0	55.0	55.0
310	54.1	53.8	53.8	54.6	54.4	54.6	55.2	55.6	55.2
311	52.7	52.9	52.8	53.4	53.5	53.4	52.9	52.9	52.8
370	54.4	54.4	54.4	54.2	54.3	54.2	53.3	53.3	53.5
375	53.2	54.1	53.0	53.4	53.4	53.2	53.0	53.4	53.2
540	53.3	53.3	53.3	52.7	52.8	52.9	53.2	52.8	52.9
571	49.5	49.6	48.9	50.4	50.1	50.7	51.0	50.6	51.0
780	54.3	54.3	54.3	54.5	54.5	54.5	54.6	54.4	54.0
799	52.3	52.1	52.7	51.0	50.7	51.0	54.6	54.3	54.3
860	52.7	52.7	52.7	52.7	52.7	52.7	52.7	52.7	52.7
920	53.8	53.8	53.8	55.0	55.0	53.8	53.8	53.8	53.8
927	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0

TABLE C-II. REFERENCE VALUES FOR CARBON MONOXIDE TEST CONCENTRATIONS USED IN COLLABORATIVE TEST, MILLIGRAMS PER CUBIC METER

Laboratory Code Number	Low Concentration	Intermediate Concentration	High Concentration
220	8.4	29.9	52.6
222	8.6	30.2	52.1
253	8.5	30.1	52.3
270	8.4	29.9	52.2
310	8.5	29.8	52.2
311	8.6	29.8	52.6
370	8.6	29.9	52.1
375	8.2	30.2	52.3
540	8.5	29.9	52.3
571	8.5	30.0	52.3
780	8.4	29.8	52.2
799	8.4	30.2	52.3
860	8.2	29.9	52.3
920	8.6	29.8	52.3
923	8.6	30.1	52.3
927	8.4	30.1	52.3

NOTATION

(a) *Principal Variables:* (may also be used as subscripts)

- y = measurements,
- L = laboratories,
- M = materials or concentrations,
- D = test days, and
- e = replication errors.

(b) *Qualifying Subscripts:*

- i = a particular laboratory,
- j = a particular material,
- k = a particular test day, and
- m = a particular replication error.

(c) *Number of Levels of Variables:*

- p = number of laboratories,
- q = number of materials,
- w = number of test days, and
- n = number of replicates.

(d) *Statistical Notation:*

- c_{ij} = the reference value for the *jth* material for the *ith* laboratory,
- $\bar{c}_j$ = the average of all c_{ij} for material j,
- $\bar{y}_{ijk}$ = average of all replicates by laboratory i on material j on day k, and
- $\bar{\bar{y}}_{ij}$ = average of all $\bar{y}_{ijk}$ by laboratory i on material j.

(e) *Measures of Variability:*

- σ = population standard deviation,
- s = sample estimate of σ,
- R = range (largest measurement minus smallest measurement), and
- $V(y)$ = variance of random variable y.

(f) *Analysis of Variance:*

- DF = degrees of freedom,
- SS = sum of squares,
- MS = mean square, and
- $E(MS)$ = expected value of mean square.

(g) *Regression Analysis:*

- x = independent variable,
- y = dependent variable,
- β = slope of a straight line,
- d = residual (observed value minus fitted value), and
- r = correlation coefficient.

(h) *Linear Model Analysis:*

- x_j = average of all $\bar{\bar{y}}_{ij}$ for material j,
- $\bar{x}$ = average of all x_j,
- α = the slope of the line β_i versus μ_i,
- β_i = slope of the line $\bar{\bar{y}}_{ij}$ versus x_j,
- γ_j = $x_j - \bar{x}$,
- δ_i = scatter of the *ith* point about the line β_i versus μ_i,
- ϵ = replication error,
- η_{ij} = scatter of the *jth* point for the *ith* laboratory about the line $\bar{\bar{y}}_{ij}$ versus x_j,
- λ = that part of η which is not accounted for by ϵ and
- μ_i = average of all $\bar{\bar{y}}_{ij}$ for laboratory i.

(i) *Qualifying Superscripts:*

- $\hat{}$ = a sample estimate of a population parameter,
- $-$ = a mean, and
- $=$ = a mean of means.

(j) *Hypothesis Testing:*

- g = number of items from which range is obtained,
- N = number of cases from which a mean is computed,
- q = a variable that has a studentized range distribution,
- s = independent estimate of standard deviation,
- t_α = the α point of a t distribution,
- z = a variable that has a normal distribution with zero mean and unit standard deviation,
- α = level of significance,
- μ = mean of a universe,
- μ_0 = hypothetical value of μ that is being tested,
- ν = degrees of freedom, and
- σ = population standard deviation.

NOTATION
Foldout for Ready Reference